Metrics and methods for quality and safety assurance of milk powder under diverse conditions

---

A thesis

submitted in partial fulfilment

of the requirements for the Degree of

Doctor of Philosphy

at

Lincoln University

by

Kizito Kene Ejeahalaka

Lincoln University

2020

Declaration

List of Publications

All aspects of the research work described in this thesis have been published in peer-reviewed journals, and they will be referred to in the text as follows:

PAPER I Kizito K Ejeahalaka, Long Cheng, Don Kulasiri, Grant R Edwards and Stephen L W On. Efficacy of near infrared spectroscopy to segregate raw milk from individual cows between herds for product innovation and traceability. ***Quality Assurance and Safety of Crops and Foods*** (Accepted and published July 2020).

PAPER II Kizito K Ejeahalaka and Stephen L W On (2019). Chemometric studies of the effects of milk fat replacement with different proportions of vegetable oils in the formulation of fat-filled milk powders: Implications for quality assurance. ***Food Chemistry***, 295, 198 – 205 (Accepted and published May 2019).

PAPER III Kizito K Ejeahalaka and Stephen L W On (2019). Effective detection and quantification of chemical adulterants in model fat-filled milk powders using NIRS and hierarchical modelling strategies. ***Food Chemistry***, 309, 125785 (Accepted and published October 2019).

PAPER IV Kizito K Ejeahalaka and Stephen L W On. Characterisation of the quality alterations in model fat-filled milk powders under inclement conditions and the prediction of the storage time using near infrared spectroscopy. ***Food Chemistry,*** 323, 126752 (Accepted and published April 2020).

PAPER V Kizito K Ejeahalaka, Peter Mclaughlin and Stephen L W On. Monitoring the composition, authenticity and quality dynamics of commercially available Nigerian fat-filled milk powders under inclement conditions using NIRS, chemometrics, packaging and microbiological parameters. ***Food Chemistry***, 339, 127844 (Accepted and published August 2020).

Abstract of a thesis submitted in partial fulfilment of the requirements for the Degree of Philosophy.

Metrics and methods for quality and safety assurance of milk powder under diverse conditions

by

Kizito Kene Ejeahalaka

Global dairy producers are moving steadily towards producing differentiated products aimed at niche markets with greater emphasis on maximising resource utilisation and meeting consumer preferences. Consequently, fat-filled milk powders (FMP) are produced by partial or whole replacement of milk fat with vegetable oils to increase the sales of underutilised skimmed milk powders (SMP), while cows with specialised characteristics and requirements may be aggregated into different herds for targeted nutritional management to reduce the milk fatty acids (FA) that are often associated with negative health effects, offering improved economic returns. Rapid methods are therefore desirable for independent verification of product quality and origin for validation and traceability of such products. We explore the use of near infrared spectroscopy (NIRS), combined with chemometrics and diverse modelling strategies, for this purpose. This study was carried out in five stages: i) the segregation of freeze-dried raw milk samples from individual cows from different herds under the same or differing feeding regimes; ii) the formulation of model FMP types according to differing protocols that reflect the differing fat sources (i.e., coconut (CO), palm (PO), soya bean (SBO) and sunflower (SFO)) used in their manufacture; iii) the adulteration of model FMP types with differing concentrations of chemical adulterants such as melamine and urea; iv) the accelerated storage of model FMP types for 7 weeks at 40 °C and 90 % RH; and v) the validation of the applicability of all the models developed for model FMP types using commercial samples covering 7 FMP brands collected from Nigerian markets.

We found that NIRS can reliably segregate freeze-dried raw milk from individual cows/herds for product innovation, providing quality assurance and traceability benefits, but may not be adequate in correlating their FA profiles in accordance with the Beer's law. The replacement of milk fat with PO, SBO and SFO resulted in FMP that were less atherogenic than whole milk powder (WMP). The PO FMP had spectral profiles and lipid indices closest to WMP but SBO and SFO FMP types offered the best indices that may support arterial health. Chemometric models developed with the NIR spectral profiles, particularly the PLS-DA models provided substantive differentiation (i.e., zero false negative and positive values with mean efficiency of 100 %) among the model FMP products. The adulteration of the model FMP types with different concentrations (i.e., 0.01 to 16.00 %) of melamine and urea resulted in some compositional alterations in the sample matrices. The NIRS profiles of CO and PO FMP types highlighted that CO had high specificity for the chemical adulterants while PO appeared to have contrasting abilities and masking their presence. The SIMCA multilevel models developed using the hierarchical modelling strategies were able to detect, confirm and differentiate the adulterations with an efficiency ranging from 89.8 to 100.0 %. The concentrations of the chemical adulterants in FMP were predicted with minimal errors ($R^2_p \geq 0.96$ and RSR ≤ 0.19) at ≥ 1.00 % adulteration level. Generally, NIRS had lower limit of detection for melamine than urea, and their concentrations were better predicted in FMP. The accelerated storage of the model FMP types for 7 weeks at 40 °C and 90 % RH significantly altered their FA, amino acids (AA) and NIR spectral profiles. The PO FMP type recorded the least spectral and FA profile alterations but the milk fat replacement with vegetable oils did not make FMP more stable than WMP. The multiclass SIMCA models provided firm differentiation of the fresh and the aged FMP samples. The PLSR models predicted the storage time with high efficiency (NSE ≥ 0.90) and low errors (RSR ≤ 0.28), indicative of FMP freshness and stability. The 7 FMP brands collected from Nigeria were of acceptable integrity and they were successfully used to validate the applicability of the qualitative and quantitative models developed in this project. This study has demonstrated the potential of NIRS as a rapid and low-cost method for robust and extensive routine authentication of dried milk powder in general and FMP in particular.

Keywords: fat-filled milk powders, NIRS, SIMCA, PLS-DA, PLSR, random selection, Kennard-Stone algorithm, iPLS, melamine, urea, vegetable oils, fatty acids, amino acids, accelerated storage, milk adulterations, chemometrics, traceability

Acknowledgements

"When one achieves massive success, unenlightened minds say, "It's all about luck." But if you were to ask me, I would say, "It is the mighty reward of faith, confidence, courage, patience, relentlessness and resiliency." I find these words from Edmond Mbiaka extremely helpful in acknowledging the generous support of many in completing this project.

First and foremost, I would like to thank the Almighty God for giving me the strength and good health to complete this project. I am immensely indebted to my supervisor, Associate Professor Stephen On, for giving me the chance to do this PhD and for providing unparalleled guidance and support throughout the project. Without a doubt, Stephen is the most supportive supervisor I have ever known, and the fate of this project would not have been the same without his invaluable contributions. I would also like to thank my associate supervisor, Professor Don Kulasiri, for his encouragement and for showing me the other ways of looking at things. A huge thanks to Dr Paul Cheng for providing the preliminary dataset to kickstart the project. My gratitude also goes to Stephen Stilwell, Dr Leo Vanhanen and the analytical team at John Burton building for providing prompt technical support during the experimental phase of this study.

I am grateful to Eurofins Laboratory Services Christchurch (c/o Peter McLaughlin) and the Clemson University Centre for Flexible Packaging (c/o Pat Guerra Marcondes) for their contributions to this research.

Finally, I would like to give a special thanks to my wife, Juliet, for being my number one cheerleader and for her unending love and support, without which this PhD program would not have been completed. Thanks to my brothers and sister for their prayers and moral support.

This work is dedicated to my son, Rex, who was born in the middle of the project, and to my sister, PCos, who suddenly passed away at the beginning of the project.

Table of Contents

List of Tables

List of Figures

Abbreviations

FMP: Fat-filled Milk Powders

WMP: Whole Milk Powder

IMP: Infant Milk Powder

SMP: Skimmed Milk Powder

CO: Coconut Oil

PO: Palm Oil

SBO: Soya Bean Oil

SFO: Sunflower Oil

FA: Fatty Acids

AA: Amino Acids

IA: Index of Atherogenicity

IT: Index of Thrombogenicity

RS: Random Selection

KS: Kennard-Stone algorithm

HMS: Hierarchical Modelling Strategies

NIRS: Near Infrared Spectroscopy

PCA: Principal Component Analysis

SIMCA: Soft Independent Modelling of Class Analogy

PLS-DA: Partial Least Square Discriminant Analysis

PLSR: Partial Least Square Regression

iPLS: Interval Partial Least Square

RSR: Root-mean-squared-error-observation Standard-deviation Ratio

NSE: Nash-Sutcliffe Efficiency

EMSC: Extended Multiplicative Signal Correction

SG: Savitzky-Golay algorithm

LOD: Limit of Detection

Chapter 1

Introduction

1.1 Research problem definition

The recent volatility in the global dairy trade has prompted producers to move steadily towards producing differentiated products aimed at niche markets with greater emphasis on maximising resource utilisation and meeting consumer preferences. For instance, cows with specialised characteristics and requirements can be aggregated into different herds for targeted nutritional management (Dooley, Parker, Blair, & Hurley, 2005) to reduce the milk fatty acids (FA) that are often associated with negative health effects, although a rapid method may be needed for segregrating their end products. However, the ideal nutritional milk fat would contain 10 % polyunsaturated fatty acids (PUFA), 8 % saturated fatty acids (SFA) and 82 % monounsaturated fatty acids (MUFA) which unfortunately cannot be accomplished by modifying the diets of lactating cows (Grummer, 1991).

Milk fat is more saturated than most plant oils (Lock & Bauman, 2004), and as such there exists the opportunity to improve the FA profile of milk by blending two or more indegeneous vegetable fats in optimal proportions with skimmed milk powder (SMP) (which usually is under-utilised and trades at a discount to whole milk powder (WMP)) to produce fat-filled milk powders (FMP) for developing markets. FMP are inexpensive recombined products formulated to resemble WMP with the substitution of milk fat by a blend of locally available vegetable oils (Early, 1988) for those in areas where milk supply is poor or even absent. A wide range of vegetable oils can be used, but their impacts and stability when blended with SMP to formulate FMP are still unknown.

FMP are packaged in bulk in more advanced, milk-surplus temperate countries to provide basic nutrition and a cheaper milk alternative for developing markets such as Nigeria, where they are shipped via sea freight containers and then repackaged into single serve/small laminated sachets for retail distribution within and across the borders of the country (Uppu, 2001). The repackaged FMP, with high fat and protein contents (28 g/100 g and 25.7 g/100 g respectively (Codex Alimentarius Commission, 1999)), are therefore exposed to harsh climatic and sanitary conditions, and a relatively long distribution and storage times (approximately

107 days), presenting challenges in preserving product quality and stability. Dried milk generally deteriorates when stored in the warehouse, retail store or on the kitchen shelf under humid conditions because of undesirable physical, chemical and/or microbiological changes, which eventually leads to a loss of nutritive value (Richards & Chandrasekhara, 1960). Moreover, as the production of FMP requires the partial or whole replacement of milk fat with edible vegetable oils, unscrupulous manufacturers could capitalize on the opportunity to tamper with the other natural constituents of the product for the purpose of increasing its apparent value or reducing the cost of its production (Handford, Campbell, & Elliott, 2016). Also, the sales for FMP are still rising (e.g. 2,266,500 metric-tons were exported globally in 2018 representing an increase of 29 % on 2013 exports (CLAL-Dairy Economic Consulting firm, 2019)), and that may provide a driver for fraudulent activity. World Health Organisation (2019) reported that about 600 million people in the world fall ill every year after eating these commonly contaminated foods. Thus, while the primary motivation for the fraudulent addition of prohibited substances into milk and milk products (see Table 1.1) is to maximise profit for the few, the impact is a real threat to public health and safety of the many.

The potential for tampering, and product deterioration in undesirable conditions is a strong driver for a rapid, inexpensive method for assessing quality and safety of dried milk powder (FMP in particular). Assessment of the provenance of such product would also be valuable in cases of fraud. It is therefore of utmost importance to develop fast, accurate, and sensitive analytical methods that will help in profiling the quality and safety of milk and milk products that are specifically produced for low-income consumers predominantly in developing countries.

Several researchers (Balabin & Smirnov, 2011; Borin, Ferrao, Mello, Maretto, & Poppi, 2006; Botros et al., 2013; Capuano, Boerrigter-Eenling, Koot, & van Ruth, 2015; Kasemsumran, Thanapase, & Kiatsoonthon, 2007; Mauer, Chernyshova, Hiatt, Deering, & Davis, 2009; Robert, Bertrand, Devaux, & Grappin, 1987; Santos, Pereira-Filho, & Rodriguez-Saona, 2013; T. Wu, Chen, Lin, & Tan, 2016; Zhang et al., 2014) have successfully used spectroscopic methods such as near infrared spectroscopy (NIRS) for the characterisation of milk constituents because they often require minimal or no sample preparation, as well as providing rapid and precise on-line analysis, with the potential for running multiple tests on a single sample (Nawrocka & Lamorska, 2013). According to Pasquini (2003), NIRS is a fast (one minute or less per sample), non-destructive, and non-invasive analytical technique with high penetration of the probing

radiation beam. In addition, NIRS has been found to be ideal for analysing large numbers of samples of a similar type and it has long path lengths which enable spectra to be measured by transmission through intact materials (Osborne, Fearn, & Hindle, 1993).

Table **1.1:** Reasons for the addition of adulterants and their methods of determination

Common milk adulterants	**Reason for adulterant addition**	**Methods for determination**	**References**
Urea	To improve the nitrogen content of the milk in order to meet its minimum protein requirement.	Potentiometric biosensor; Raman chemical imaging; gas chromatography-mass spectrometry	Ma et al. (2016); Zhang et al. (2014); Qin, Chao, and Kim (2013).
Melamine (nitrogen-rich organic compound)	Artificially stimulates increased protein content since it contains 67% nitrogen per mass unit. It increases the pseudo protein content of the milk.	Raman spectroscopy; near infrared spectroscopy; gas chromatography mass spectrometry; enzyme-linked immunosorbent assay (ELISA)	Zhang et al. (2014); Koh et al. (2011); Balabin and Smirnov (2011); Y. Cheng et al. (2010).
Starch (amylose and amylopectin)	To increase the total solids and viscosity of the diluted milk.	Near infrared spectroscopy	Zhang et al. (2014); Borin et al. (2006).
Whey (possible adulterant of both fluid and dry milk)	To increase the solid content of the milk	Chromatography or electrophoresis; near infrared spectroscopy	Zhang et al. (2014); Borin et al. (2006).
Sucrose (possibly from cane sugar or sugar beet addition)	To increase the carbohydrate content of the diluted milk	HPLC, detection by refractive index; near infrared spectroscopy	Zhang et al. (2014); Borin et al. (2006).

Thus, NIRS is a simple but versatile approach with negligible operator-induced sources of error (Osborne et al., 1993) and for all these reasons, it has been chosen for this research project. It has already been used for detecting multiple adulterants in kudzu starch (Xu, Shi, Cai, Zhong, & Tu, 2015), Chinese lotus root powder (Xu, Shi, Ye, Yan, & Yu, 2013), olive oil (Christy, Kasemsumran, Du, & Ozaki, 2004), and beef hamburger (Ding & Xu, 2000), and for monitoring the quality loss of shiitake mushroom (Xu, Fu, Cai, & She, 2015), minced beef (Sinelli, Limbo, Torri, Di Egidio, & Casiraghi, 2010) and pasteurised skim milk (Al-Qadiri, Lin, Al-Holy, Cavinato, & Rasco, 2008) during storage. It has also been successfully used for discriminating raw milk according to feeding regimes (Coppa et al., 2012; Valenti et al., 2013), and for predicting the

fatty acid composition of cow milk derived from different feeding regimes (Coppa et al., 2010). This thesis extends our knowledge, understanding, application and analysis of NIRS for complex food analysis.

1.2 Research justifications

Although NIRS has been extensively applied in milk research, our literature review revealed that there are still gaps and opportunities (Table 1.2) to refine it as a method, characterise NIR spectral relationships with dairy product composition and variance and explore its application to a poorly studied issue for developing nations.

Table **1.2:** Gaps in existing knowledge base requiring further research

Existing knowledge base	Gaps in existing knowledge base
NIRS has been used in a multi-factorial research to discriminate raw milk from different regions according to feeding regimes and fatty acid profiles (Coppa et al., 2010; Coppa et al., 2012; Valenti et al., 2013).	It has never been applied to discriminate raw milk from New Zealand and to segregate those from individual cows in the same herd, at the same location, and fed on the same or different regimes.
NIRS has generally been applied to milk products (Balabin & Smirnov, 2011; Botros et al., 2013; Capuano et al., 2015; Kasemsumran et al., 2007; Mauer et al., 2009; Santos et al., 2013; T. Wu et al., 2016; Zhang et al., 2014).	It has never been applied to FMP to the best of our knowledge. Consequently, a method is needed since NIRS calibrations are sample matrix dependent (Osborne et al., 1993).
NIRS has been used to monitor quality loss of pasteurised skim milk during storage (Al-Qadiri et al., 2008).	It has never been used to monitor the behaviour of FMP types over time under the conditions of environmental stress.
NIRS has been used to explore the sources of variance in authentic bovine skim and nonfat milk powders from the USA, New Zealand, Ireland and Denmark (Botros et al., 2013).	It has never been used to explore the variances in FMP formulated with different backbone fats and in those that are commercially available. The regional milk source variability could have a major influence on the composition of FMP brands exported from different countries.
NIRS has been successfully used to detect and quantify chemical adulterants such as melamine in milk powders (Balabin & Smirnov, 2011; Schoder, 2010; Zhang et al., 2014).	It has never been used to screen for chemical adulterants in FMP.

1.3 Aims and objectives

The overall aim of this study is to examine, in detail, the use of near infrared (NIR) analysis for quality and safety assessments of milk and milk products, notably relating to those available

in developing countries such as Nigeria. To do this, understanding NIR spectral variation at micro- to macro levels is essential, to place results into context and so make for accurate conclusions. This will be achieved in the following objectives:

1. Assessment of spectral variation in individualised raw milk samples of New Zealand origin and relation to herd, and detailed fatty acid composition using the previously generated body of analysis and data (Paper I).

2. Assessment of the potential of NIR spectral profiling to discriminate FMP formulated with different backbone fats (Paper II).

3. Evaluation of the specificity and sensitivity of NIR spectral analysis in detecting chemicals used to adulterate FMP products (Paper III).

4. Assessment of NIR spectral profile shifts in FMP stored under conditions relating to the Nigerian environment (high temperature, high humidity), to evaluate its potential to characterise degradation (Paper IV-V).

5. Application of NIR spectral analysis for the discrimination of FMP brands sourced from different companies and trading routes/commercial centres (Paper V).

6. Use of method to examine samples from a developing country such as Nigeria, to assess the quality and safety of FMP (Paper V).

7. Establishment of the background levels for key microbial flora to benchmark possible variance of potentially adulterated samples (Paper V).

1.4 Research questions

The research questions to be addressed in the course of this study are as follows:

1. To what extent can NIR spectral analysis be related to fatty acid profiles of individual raw milk samples, and discriminate milk from different cow herds?

2. Can NIR spectral analysis be used to differentiate FMP products with differing backbone fats used in their manufacture?

3. Can NIR profiling be used to determine the quality of FMP samples stored under conditions of environmental stress?

4. How reliable is NIR analysis in discriminating FMP products originating from different countries and brands?

5. What is the efficacy (specificity, sensitivity) of NIR spectral profiling for detection of adulterants, and what conclusions can be drawn from the examination of FMP products from Nigeria?

6. What relationship, if any, is there between microbial content of samples and potential adulterants?

1.5 Our approach

In this study, the NIRS measurements will be completed using the FOSS NIRS DS2500 F analyser. Our studies will examine sample variation at micro-to-macro level (geographic and compositional), combined with appropriately complex analyses, to enhance our understanding of NIRS applications and to fill important data gaps in the field.

Investigations will commence with the study of the NIR spectroscopic variation of raw milk from individual cows across four herds in New Zealand, for which extensive information on herd location, feeding regime and fatty acid composition is known. Šašić and Ozaki (2001b) showed in their study that fat was the main source of spectral variances in raw milk in the short-wave NIR region. The NIR absorption spectrum of milk was found (Frankhuizen, 2001) to be very similar to that of water, except that milk contained scattering particles in the form of fat globules and protein micelles which intensified the absorption bands. Representative samples from each of the herds will be objectively selected using mathematical methods. The spectral data of these representative freeze-dried raw milk samples will be examined to explore the resolution of samples to herd level and to their fatty acid profile, and the outcome will be used to establish a high-quality baseline standard for all future comparisons involving diverse samples of FMP. Additional mathematical approaches will be used to explore the potential of linking the NIR spectra with key FA components in the milk sample. We will investigate the potential for objectively distinguishing the spectral variances in FMP using product from New Zealand modified onsite to commercial specifications, and according to differing protocols that reflect the differing fat sources (e.g. palm, coconut, soya-bean or sunflower oil) used to manufacture FMP. We aim then to progress with storage stability trials in a climatic chamber in order to understand the spectroscopic behaviour of FMP over time under diverse conditions of environmental stress typical of tropical countries, the export

destination for the milk powders. The motives are to document with certainty the fingerprint of FMP of genuine quality and to gain valuable insight into its long-term behavior when exported to different countries under different environmental conditions. We will then examine the spectra of FMP with different quantities of known adulterants (e.g. urea, melamine) (Qin et al., 2013) in order to evaluate the efficacy of NIRS to readily identify such chemicals in FMP circulating in developing countries. We will conclude the study with the profiling of unknown commercial FMP samples, from different brands and trading routes in Nigeria, using a suitable prediction model for accurate quantification of adulterants and contaminants, if any, for quality and safety assurance. The microbiological (i.e. aerobic plate count, sulphite-reducing clostridia spores count, *Salmonella* and *Cronobacter sakazakii*) characteristics of the FMP samples will be documented in products sourced from Nigeria to examine for the presence of microbial contamination and any correlation with the levels of possible adulteration.

The outcomes of the research project are expected to provide crucial insights into: 1) the use of NIRS for characterising micro- to- macro variations necessary for producing differentiated raw milk for the production of high value niche products and for determining provenance of milk products for human health and safety; 2) the application of NIRS on FMP (a commodity that has never before studied) for quality and safety assurance; 3) the baseline data for FMP quality authentication and assurance, thereby simulating further research in the subject area; 4) the stability of FMP products exported to tropical developing markets and the early prediction of storage time to minimise, mitigate or prevent potential impacts of environment stresses; 5) the quality dynamics and safety of FMP brands that are commercially sold in Nigeria and as a result influencing the quality of life of millions of resource-poor people in the West African sub-region that depend on this dried milk powder for their essential nutrients.

This research project has been structured into eight chapters with five of them (i.e., chapters 3 to 7) presenting the core investigations that were undertaken. These five chapters are written to be stand-alone research articles and they have been submitted for publication in reputable international journals with four of them already published and the other one undergoing the peer-reviewed process (see page iv for details). Hence, they will be reported here in the format of the journals to which they were submitted. The first chapter defines the research problem and provides a general introduction to the research. Chapter two describes the key concepts, theories and models that underpin the research project while chapter eight presents an overall conclusion from the study.

Chapter 2

Theoretical background

2.1 Key research concepts

This section discusses the key concepts in the research based on a literature review to better understand the aims and background of the project.

2.1.1 Milk fat alterations

There is increasing evidence to indicate that excessive consumption of medium-chain SFA and trans-fatty acids (TFA) are risk factors for cardiovascular disease, and in the aetiology of other chronic conditions (Shingfield, Bonnet, & Scollan, 2013). Hence, altering the FA composition of ruminant-derived foods (e.g., milk), which are significant sources of medium-chain SFA and TFA, offers the opportunity to align the consumption of FA in human populations with public health policies without the need for substantial changes in eating habits (Shingfield et al., 2013). Milk fat may be altered through the cow's diets or replaced with a blend of vegetable oils to reduce the high risk components or to enhance the microcomponents that may have beneficial effects on health maintenance and disease prevention (Lock & Bauman, 2004). In this sense, making targeted modifications to the FA profile of milk has the potential to significantly contribute to the production of dairy products with higher added value (Hanuš, Samková, Křížová, Hasoňová, & Kala, 2018).

The opportunity to alter the FA composition of milk through nutrition of the dairy cows has improved as a result of recent advances that have better defined the interrelationships between rumen fermentation, lipid metabolism, and milk fat synthesis (Lock & Bauman, 2004). When dietary material enters the rumen, it essentially enters a large fermentation vat where the lipids are extensively altered resulting in marked differences between the FA profile of lipids in the diet (mostly unsaturated fatty acids (UFA)) and lipids leaving the rumen (mostly SFA) (Jenkins, Wallace, Moate, & Mosley, 2008). The fate of dietary lipids in the rumen depends on their release from the feed, their adherence to or incorporation in bacterials cells, their biohydrogenation (BH) and their passage to the lower gut (Elgersma, 2015). The ruminal microbes transform the lipids entering the rumen via 2 major processes, lipolysis (i.e., the transformation of the lipids entering the rumen by microbial lipases causing the release of the FA) and BH (i.e., the conversion of UFA to SFA by ruminal microbes followed by hydrogenation

of the double bonds) (Jenkins et al., 2008). Lipolysis is a prerequisite of rumen BH and it is the hydrolysis (release) of FA from the glycerol backbone of triacylglycerols, galactolipids or phospholipids (Fievez, Vlaeminck, Jenkins, Enjalbert, & Doreau, 2007). A reduced degree of lipolysis is predicted when the retention time of lipid in the rumen decreases (Elgersma, 2015). The process of BH reduces the rumen outflow of PUFA and contributes to the accumulation of *cis* and *trans* isomers in ruminant products, including conjugated linoleic acid (CLA) and *trans* monoenes (Fievez et al., 2007). Hence, the extent and type of the rumen BH process will determine both the amounts and structures of FA leaving the rumen (Fievez et al., 2007). However, the amount and quality of herbage consumed and the concentration and fate of FA in feed, digestive tract, and mammary gland determine the FA profile in dairy products (Elgersma, 2015). Cows that are pasture-fed have been found (Rugoho, Liu, & Dewhurst, 2014) to have milk FA composition that was characterised with high levels of α-linolenic acid (ALA, C18:3) and CLA, and with low levels of linoleic acid (LA, C18:2). The concentrations of these FA in raw milk were markedly different (Rugoho et al., 2014) when the pasture component of the cow diet was gradually substituted with other feeding/nutrient sources. Hence, the milk from grazed ruminants may have higher levels of PUFA than from ruminants that are kept indoors (Elgersma, 2015).

The FA composition of milk can also be altered by replacing milk fat with a blend of vegetable fats, although such replacement may well occur illegitimately (Ntakatsane, Liu, & Zhou, 2013) to defraud consumers. Consequently, FMP contain vegetable fat, in place of milk fat, together with non-fat milk solids (Gordon, 1997). FMP are formulated by blending SMP and vegetable fat, to which emusifiers, stabilisers, flavouring and colouring agents may also be added (Schmidmeier, O'Gorman, Drapala, Waldron, & O'Mahony, 2019). Thus for FMP, the vegetable oils are the dominant lipidic ingredients while the proteins and carbohydrates from the SMP represent the encapsulating matrix. Melting point is the most important factor when choosing the fat type, but the surface tension and possible content of impurities are other factors to be considered (Hansen, 1980). The following unit processes are involved in the manufacture of FMP: a) fat separation (i.e., the separation of raw milk into skim milk and cream); b) preheating (i.e., heating of the skim milk between 75 and 120 °C to destroy bacteria, inactivate enzymes, delay the onset of oxidised flavour, and impact heat stability); c) evaporation (i.e., boiling the preheated milk to remove excess water and to concentrate it to solids content of about 45 %); d) mixing and standardising the fat level (i.e., mixing the skimmed milk concentrate with the

appropriate amount of melted vegetable fat and other ingredients and homogenising at high pressure under controlled conditions (i.e., for effective dispersion of the fat) to produce a stable emulsion in which most of the fat globules are less than 1 mm in diameter (Gordon, 1997)); e) drying (i.e., drying the oil-in-water emulsion at conditions chosen carefully to minimise the free fat levels, and the dried powder is cooled as rapidly as possible (Gordon, 1997)); f) instantising; and g) packaging and storage. Table 2.1 (Codex Alimentarius Commission, 1999) shows the comparison of the nutritive values of typical FMP and the regular milk powder

Table **2.1:** Comparison of nutritive value between filled milk powder and regular milk powder

Nutrient	Filled milk powder (per 100g)	Regular milk powder (per 100g)
Energy (kcal)	505.0	506.0
Fat (g)	28.0	28.2
Protein (g)	25.7	25.7
Carbohydrate (g)	37.6	37.4
Mineral (g)	5.7	5.7
Sodium (mg)	350.0	350.0
Calcium (mg)	930.0	830.0
Water (g)	3.0	3.0
Vitamin A (IU)	1800	1800
Vitamin C (mg)	30	30
Iron (mg)	10	10

Source: Codex Alimentarius Comission (1999)

2.1.2 Milk powder storage modifications

The US industry standard for shelf-life of WMP is 6 to 9 months under storage at < 27 °C and < 65 % relative humidity, although previous research has demonstrated that flavour changes by 3 months at ambient storage (Lloyd, Hess, & Drake, 2009). High lipid milk powders with added plant oils or lipids rich in polyunsaturated, long chain FA have lower storage stability (H. Cheng et al., 2017) especially in hot countries where the product may be subjected to high temperature (above 40 °C) at some stage during their transport and storage (Hurrell, Finot, & Ford, 1983). The storage of milk powder under unfavourable conditions accelerates the normally slow deterioration in nutritional quality (Ford, Hurrell, & Finot, 1983). According to Stapelfeldt, Nielsen, and Skibsted (1997), the combination of high temperature and high water activity should not be used for accelerated storage of WMP with the sole purpose of evaluating the oxidative stability, but it is a combination worth trying for the assessment of the overall milk powder resistance to tropical storage conditions, as these promote both autoxidation, non-enzymic browning and changes of functional properties such as solubility.

Higher temperature during storage of food powders are known to accerate quality deterioration as both physical transformations and chemical reactions in most cases increase their rate with increasing temperature (H. Cheng et al., 2017). In fact, the processes of oxidative deterioration are accelerated by a factor of ten for a 10 °C increase in temperature (Stapelfeldt et al., 1997). The shelf-life of WMP clearly depends on the preheat treatment of the milk (with medium to high-heat preferred), the water activity of the product (which correlates with the rate of autoxidation) and the storage temperature (Stapelfeldt et al., 1997). For accerelated storage stability testing of powders, the temperatures are recommended not to exceed 55 °C (H. Cheng et al., 2017).

Lipid oxidation, lactose crystallisation and Maillard reactions (non-enzymatic browning) (which are driven by temperature and moisture content) are the main problems that arise over time when milk powders are shipped and stored around the world for consumer use (Chopovda et al., 2016). The oxidation of lipids is one of the major causes of quality deterioration in food, which results in the changes of nutrition, taste, texture, and appearance, and may even shorten the shelf life of food and limit its use as beneficial lipids in functional food (Sun, Wang, Chen, & Li, 2011). The rate of lipid oxidation is highly dependent on water activity, a_w, and it reaches a minimum at about $0.2 < a_w < 0.4$, with lower values of a_w targeted in hot humid regions (Chopovda et al., 2016). Lipids with PUFA content are susceptible to oxidation and the development of undesirable flavours and toxic products during storage, making their use in foods a challenge for the food industry (Sun et al., 2011). Oxygen interacting with unsaturated lipids through autoxidation is an important reaction because it limits their application in functional food (Sun et al., 2011). The oxidation of an unsaturated lipid in a heterogeneous system (i.e., oil-in-water emulsion) is different from that in a homogeneous system (i.e., bulk oil) (Sun et al., 2011). The behaviour of oil-in-water emulsions in food is determined by the three regions of the system – the oil that is in the interiors of the emulsion droplets, the interficial material between this lipid material and the aqueous phase, and the aqueous phase itself (Sun et al., 2011).

Lipids containing UFA undergo spontaneous peroxidation triggered by free radical species such as peroxyl radicals and non-radical species such as singlet oxygen (Sun et al., 2011). The formation of free radicals is the initial step in lipid oxidation (Stapelfeldt et al., 1997). It involves the loss of a hydrogen radical in the presence of trace metals, light or heat (Frankel, 1984). The abstraction of hydrogen atoms from allylic carbon atoms in unsaturated lipids

generates a lipid radical (each free radical has the potential to initiate the chain reactions of the autoxidation cycle) and breaks the spin barrier for reaction between ground state oxygen and the lipids (Stapelfeldt et al., 1997). In the propagation process, the peroxy radicals react with more unsaturated lipids to form lipid hydroperoxides which are the fundamental primary products of autoxidation (Frankel, 1984). The decomposition of lipid hydroperoxides constitutes a very complicated process and produces a multitude of materials that may have biological effects and cause flavour deterioration in fat-containing foods (Frankel, 1984). The secondary lipid oxidation products, such as the aldehydes, formed by the cleavage of the lipid hydroxides (i.e., the primary lipid oxidation products) have a profund impact on both sensory and functional properties (Stapelfeldt et al., 1997). The development of fat rancidity in complex food systems is also greatly affected by the interactions of proteins and amino acids with lipid oxidation products (Frankel, 1984). Proteins, peptides and amino acids are also susceptible to oxidative changes caused by free radicals that are manifested to a greater or lesser degree in different types of milk, depending on the quantity and type of lipids, prooxidants, antioxidants and storage conditions (Scheidegger et al., 2013). Oxidation can occur at both the protein backbone and on the amino acid side-chains, with the ratio of attack dependent on a number of factors (Davies, 2005). The ease of oxidation of sulfur centres makes cystine and methionine residues major sites of oxidation within proteins (Davies, 2005). Such oxidation can arise either via direct interaction with the oxidant, or via radical transfer processes within a protein after oxidation at remote sites (Davies, 2005). The protein oxidative status measured as the protein carbonyl content was slightly lower in SMP (total fat = 0.1 g/100 mL) than in WMP (total fat = 3.0 g/100 mL), whereas that for infant milk powders (IMP) (total fat = 3.4 g/100 mL) was more than twice as much as the value recorded for WMP (Scheidegger et al., 2013).

The greatest threat to the storage stability of products with high fat content is oxidation, and its effects can be considerably reduced by cold storage and by the addition of antioxidants (Renner, 1988). Antioxidants such as tocopherols, several amino acids and proteins, and ascorbic acid, inhibit lipid oxidation when present in foods like FMP and infant formulas, thereby increasing their shelf-lives and allowing them to be transported and stored for long periods (Angulo, Romera, Ramirez, & Gil, 1998). However, the stability of the emulsion depends on the protein concentration and amphiphilicity (Thomas, Scher, Desobry-Banon, & Desobry, 2004).

Lactose crystallisation occurs once the temperature of the powder increases above the glass transition temperature, until it reaches the melting point of the powder (Chopovda et al., 2016). Thus, moisture permeation over time through the bag packaging increases the water activity, lowering the glass transition temperature and increasing the dangers of lactose crystallisation and Maillard reactions (Chopovda et al., 2016). Proteins are partly hydrophilic and as such may delay lactose crystallisation because of their competitive water sorption (i.e., the water sorbed by milk protein may not be available for lactose crystallisation) (Thomas et al., 2004). On the other hand, lactose crystallisation is delayed by the presence of milk fat since lipids are hydrophobic (Thomas et al., 2004). Hence, fat may act as hydrophobic barrier and may both limit the diffusion of hydrophilic molecules and the growth of lactose crystals (Thomas et al., 2004). Nevertheless, FMP contain a mixture of vegetable oils with different melting points and as such the fat with lower melting point can be released onto the surface of the powder more easily resulting in fat bridging and cake formation during storage (Tham, Xu, Yeoh, & Zhou, 2017).

Maillard reaction is a chemical reaction between reducing sugars such as lactose and proteins (and breakdown products of proteins such as peptides) (Renner, 1988). The extent of the Maillard reaction depends not only on temperature and moisture content, but also on storage time and the composition of the product (Hurrell et al., 1983). During prolonged storage of milk powders, Maillard reactions between milk proteins and lactose may reduce the level of the essential amino acids, particulary that of lysine (Hurrell et al., 1983).

2.1.3 Milk powder adulterations

Food safety (i.e., unintentional contamination of food causing unintentional harm), and food fraud (i.e., intentional substitution, addition, tampering, or misrepresentation of food for economic gain) incidents can create adulteration of food with public health threats (Spink & Moyer, 2011). The U.S Federal Food, Drug, and Cosmetic Act Section 342 defines adulterated food principally as food that bears or contains: “any poisonous or deleterious substance which may render it injurious to health; but in case the substance is not an added substance such food shall not be considered adulterated under this clause if the quantity of such substance in such food does not ordinarily render it injurious to health” (Manning & Soon, 2016). Milk products are of high nutritional value and the increased demand has made them prone to fraudulent activity (Handford et al., 2016). A review of 660 scholarly, media, and other publically available reports of the top 25 most frequently adulterated food ingredients

worldwide from 1980 to 2010 showed that milk contributed 14 % of the records and was only second to olive oil which covered 16 % (Moore, Spink, & Lipp, 2012). Milk may be adulterated with complex organic compounds such as melamine or urea, or even with simple substances such as water for financial gain. The practice of adulteration of milk invariably reduces its quality and may introduce hazardous substances into the dairy supply chain jeoparding consumers' health (Poonia et al., 2017).

Melamine is a white, crystalline, triazine compound (1,3,5-triazine-2,4,6-triamine) that has a high nitrogen content (66 % by mass) and is used in making thermo-setting resins, as a fire retardant (due to its release of nitrogen gas when burned) and as a pigment (Smith, Inomata, & Peters, 2013). Melamine is manufactured by the decomposition of urea to cyanic acid and ammonia, followed by a condensation reaction of the cyanic acid (Smith et al., 2013). Urea, on the other hand, is an end product of nitrogen metabolism and a normal constituent of milk at levels between 180 and 400 ppm, making up about 55 % of the non-protein nitrogen compounds (Hilding-Ohlsson, Fauerbach, Sacco, Bonetto, & Cortón, 2012). The presence of urea in milk above the cut off concentration of 700 ppm could cause severe health problems for humans (Hilding-Ohlsson et al., 2012). Unfortunately, melamine and urea are frequently added to milk to elevate its nitrogen content and mimic a high protein concentration (Hilding-Ohlsson et al., 2012). The driving force for this adulteration is that the protein content of milk products is traditionally assessed through non-specific procedures such as the Kjeldahl reaction and the Dumas method. These procedures quantify the nitrogen contents of amino acids and therefore of proteins, but lacked the ability to distinguish them from those of many other non-proteinaceous molecules such as melamine and urea (Abernethy, Sheehan, Griffiths, & Williams, 2009). Consequently, unscrupulous manufacturers capitalise on the lack of specificity of these procedures for nitrogen to add nitrogen-rich chemicals into milk powders to spuriously inflate their apparent protein contents. This therefore enabled them to dilute their milk and still maintain admissible protein-level readings (Everstine, Spink, & Kennedy, 2013).

In 2008, high levels (up to 2,500 ppm; acceptable threshold = 1 ppm) of melamine were detected in some IMP produced in China affecting over 290,000 infants and causing six deaths (Hilding-Ohlsson et al., 2012). The maximum amount of melamine allowed in IMP is 1 mg/kg (1 ppm or 0.0001 % w/w) and the amount of the chemical allowed in other foods and animal feed is 2.5 mg/kg (2.5 ppm or 0.00025 % w/w) (FAO, 2010). The ingestion of high melamine

concentrations is particularly harmful because of the formation of a very stable complex with cyanuric acid which crystallises causing renal damage (Finete, Gouvêa, de Carvalho Marques, & Netto, 2013).

Due to the serious health concerns associated with melamine consumption, rapid and sensitive methods to detect melamine's presence are essential (Mauer et al., 2009). High performance liquid chematography and infrared spectroscopy were the most common analytical detection procedures, and chemometrics data analysis was used in a large number of the 660 reports of the top 25 most frequently adulterated food ingredients worldwide from 1980 to 2010 reviewed by Moore et al. (2012).

The study by Schoder (2010) demonstrated that, although clinical cases of melamine poisoning were still rare in Africa due to the lack of complex ultrasonographic equipment and expertise required for its diagnosis, the contamination of milk powder with melamine is a significant problem in the continient. The study showed that 6 % and up to 11 % of random samples tested for melamine in the Dar-es-Salaam region of Tanzania were positive and the health problems caused by the contamination were likely to be substantial (Schoder, 2010).

At least two general analytical strategies to detect food adulteration have evolved (Moore, DeVries, Lipp, Griffiths, & Abernethy, 2010): 1) using analytical tests to identify one or more suspected adulterants (known a priori), especially at low levels, with the absence of result indicating that the test material was not adulterated with a specific material; 2) using compendial identification and purity tests to substantiate the ingredient's identity/purity and, as a result, identify known or unknown adulterant that was substituted for the original material at concentrations sufficiently high to be recognised by perturbations in the test results. This approach remains a useful tool for reducing the risk of adulteration since the counterfeiters must adulterate at high relative concentration levels to realise economic gain (Moore et al., 2010).

2.1.4 Near infrared spectroscopy

Near infrared spectroscopy (NIRS) is based on the absorption of electromagnetic radiation at wavelengths in the range 780 – 2500 nm (Fig. 2.1) (Osborne et al., 1993). The NIR spectrum originates from radiation energy transferred to mechanical energy associated with the motion of atoms held together by chemical bonds in a molecule (Pasquini, 2003). When a sample is irradiated with light, fractions are reflected, transmitted and absorbed all summing to 1.0,

with the proportions depending on the light wavelength and the sample properties such as composition, thickness etc (Agelet & Hurburgh, 2010). That implies that in a given wavelength range, some frequencies will be absorbed, others (that do not match any of the energy differences possible for that molecule) will not be absorbed while some will be partially absorbed (Pasquini, 2003). The absorbance recorded in the NIR region is due to combinations and overtones of the fundamental molecular bond vibrations as discussed later in this section. The Beer's law correlates the sample concentration with its absorbance at specific wavelengths (Agelet & Hurburgh, 2010). The higher the concentration of the sample, the higher the absorbance values and the lower the reflectance. The wavelengths at which the sample absorbs light can be measured using a near infrared spectrophotometer and the results obtained can be plotted as shown in Fig.2.2 in order to determine some basic properties of the sample. Peak absorbances occur at wavelengths (referred to as the maximum or optimal wavelengths) at which the sample is most sensitive to changes in concentration. The wavelength positions of absorbance bands are specific to the functional groups in a sample and as such each sample has a unique "fingerprint" absorbance spectrum (Mauer et al., 2009). Thus, the successful implementation of NIRS will require matrix-specific validation and the use of supporting reference materials (Moore et al., 2010). Hence, this method may be suitable only for food ingredients that have a low degree of compositional variability and are not complex finished food products (Moore et al., 2010). In general, it is only suitable for measuring materials that can absorb infrared light (i.e., water, protein, fat, starch, sugar etc.).

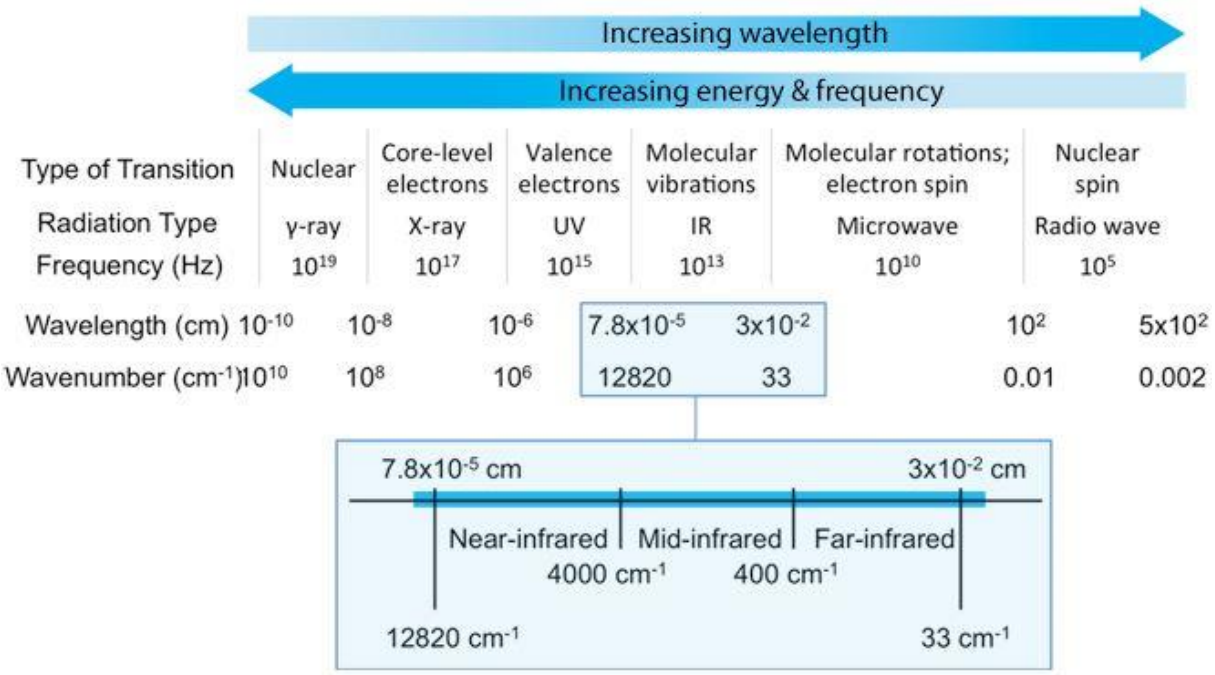


Figure **2.1:** The electromagnetic spectrum and the infrared region
Source: www.chroacademy.com

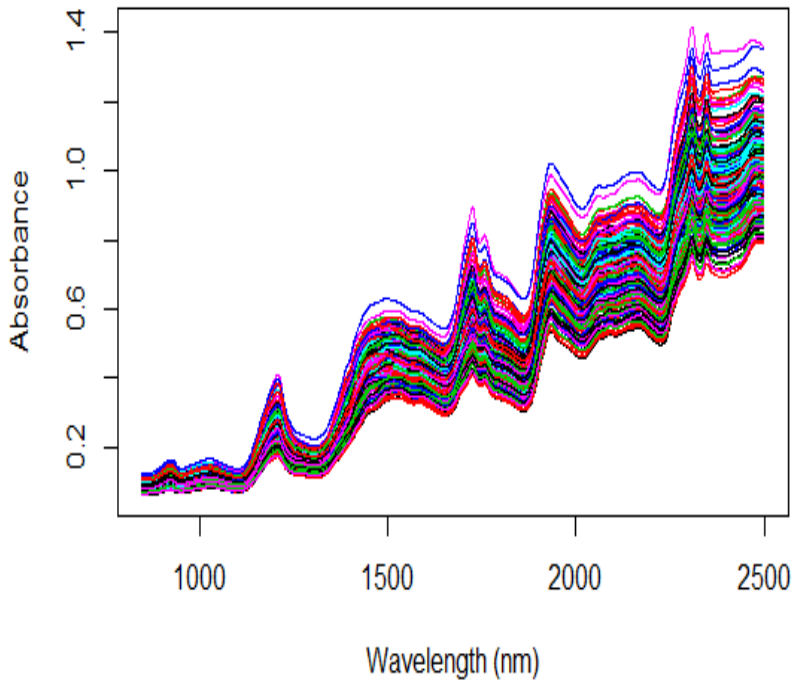


Figure **2.2:** Typical raw near infrared spectra of freeze-dried fat-filled milk powder samples (Ejeahalaka & On, 2019a)

The NIR spectra of foods comprise broad bands arising from overlapping absorptions corresponding mainly to overtones and combinations of vibrational modes involving C—H, O—H and N—H chemical bonds (Osborne, 2006). Overtones are electron excitations to higher energy levels while the combination bands, which are located at higher NIR wavelengths (i.e., 1900 to 2500 nm), basically involves a combination of vibrations from the same chemical groups of the overtones, but as a result of the interactions between molecular vibrational frequencies, overlapped information from Fermi resonances etc (Agelet & Hurburgh, 2010).

Rapid infrared spectroscopic methods such as NIRS have been successfully used in a wide range of complex investigations for profiling, classification, quantification and for adulteration detection of milk powders (Mauer et al., 2009). NIRS is used routinely for the compositional, functional and sensory analysis of food ingredients, process intermediates and final products (Osborne, 2006). According to Moore et al. (2012), high-performance liquid chromatography and infrared spectroscopy were the most common analytical detection procedures while chemometrics was the most used data analysis technique in a large number of the 660 reports of the top 25 most frequently adulterated food ingredients worldwide reviewed from 1980 to 2010. However, the traditional or conventional methods (see Section 2.2) of analysis are currently being replaced by spectroscopic techniques such as NIRS because of the following advantages (García-Sánchez, Galvez-Sola, Martínez- Nicolás, Muelas-Domingo, & Nieves, 2017): i) they are non-destructive and do not polute the environment since they do not use chemical reagents; ii) they require minimal or no sample preparation and as such the analyses are very simple, cheap and fast; iii) they measure many parameters in a single analysis (i.e., multi-parametric); and iv) they can perform analyses *in situ* and online for a large number of samples per minute. The major limitation of NIRS in food analysis is its dependence on less-precise reference methods (Osborne, 2006). Milk powder is one of the most suitable products to be analysed with NIRS because of its uniform particle size and shape (Chen, Tan, Lin, & Wu, 2018).

2.2 Conventional milk powder investigations

In practice, the discrimination of milk powder is conventionally carried out by visual inspection, but this method is relatively subjective (Chen et al., 2018). An objective and reliable assessment of milk samples often involves the use of complex techniques such as gas chromatography, high performance liquid chromatography etc. which are time consuming, laborious, expensive and almost impossible to be enforced conveniently (Chen et al., 2018).

Similarly, Kjeldahl and Dumas (combustion) methods are widely accepted for total protein determination but they lack analytical selectivity for protein because they measure protein on the basis of sample nitrogen content (Moore et al., 2010). Adulteration incidents exploiting this analytical vulnerability (for example, melamine) demonstrate that these methods are no longer sufficient to protect the public health since they are not protein-specific (Moore et al., 2010). Alternative conventional methods such as liquid chromatography—triple-quadrupole tandem mass spectrometry (LC-MS/MS) can detect melamine residues in IMP with limits as low as 0.25 ppm but the sample preparation and clean-up procedures are time-consuming and labour-intensive (Mauer et al., 2009). Therefore, the use of this method as a screening tool for a large number of samples is not practical or cost-effective (Mauer et al., 2009). In addition, a rapid method using a low-temperature plasma probe in combination with tendem mass spectrometry (LTP-MS/MS) has a detection limit of 0.006 ppm in milk powder but the equipment may not be widely available or exportable for international adoption of the method (Mauer et al., 2009). These limitations suggest that there is still a need for a rapid, high-throughput, widely available, cost-effective method for detecting melamine in milk powder (Mauer et al., 2009).

However, although primary methods may be viewed by many as old fashioned, the results from validated chemical methods performed on a small number of samples form the basis for calibration of the high-speed testing methods (Barbano & Lynch, 2006). Thus, the accuracy of these primary or reference tests is of paramount importance because all NIRS milk testing is dependent on it (Barbano & Lynch, 2006).

2.3 Chemometric modelling of milk powder functionalities using NIRS

Chemometrics enable the interpretation of molecular structural information and correlation to sample composition by processing and interpreting complex spectra using multivariate statistical tools (Mauer et al., 2009). They allow the correlation of subtle spectral variability with the sample composition and thus yields effective chemical content detection and quantification (Beć & Huck, 2019). Analytical NIRS relies extensively on various methods of statistical analysis (called chemometrics) which may be roughly divided into three classes (Beć & Huck, 2019): 1) exploratory data analysis using data mining techniques such as the principal component analysis (PCA) to identify any patterns and characteristics in order to gain deeper insights into the large set of NIR spectra; 2) classification analysis using approches such as soft independent modelling of class analogy (SIMCA), partial least square discriminant analysis

(PLS-DA) etc to separate, sort and group samples in accordance with their source of origin, level of authenticity or even their storage time; 3) regression analysis using techniques such as partial least square regression (PLSR) to predict/quantify chemical content. NIR spectroscopic models are not matrix independent and as such new calibrations are required for application to different brands or formulations of the same milk powder (Mauer et al., 2009).

2.3.1 Spectral collection

The data collection process begins with a classical analysis of the desired components in each of the samples to obtain the reference values and then the NIR measurements to record their spectra (García-Sánchez et al., 2017). The most common measurement modes employed by NIRS include (Pasquini, 2003): i) transmittance (i.e., log(1/T); where the measurements (e.g., for liquids, gaseous and translucent samples) are done with typical optical path); ii) transflectance (i.e., where the measurements (e.g., for thin or clear samples) are done by doubling the optical path as the radiation beam passes twice through the sample); and iii) diffuse reflectance (i.e., log(1/R); where scattering and absorbance by solid granules contribute to change in the signal intensity. It is used generally for solids and other light scattering matrices). In NIR diffuse reflectance, the incident beam is scattered within the sample, and returned to the surface after within-sample absorption (Agelet & Hurburgh, 2010). Diffuse reflectance mode (best measured between 1200 and 2500 nm) allows working with thicker and denser samples without inducing as much heating as transmission (which favours measurements at lower wavelengths) (Agelet & Hurburgh, 2010). The spectral data collected from several scans of the samples (i.e., milk powders) will first be visualised to establish whether there are common relationships, patterns, trends, noisy areas, outliers or even quality impairments that could help outline the analysis. Then, the data will be pre-processed to remove physical phenomena in the spectra in order to improve the subsequent exploratory analysis, classification modelling and multivariate regression (Rinnan, Van Den Berg, & Engelsen, 2009).

2.3.2 Spectral pre-processing

The NIR spectral data normally undergo pretreatment mainly to overcome the problems associated with radiation scattering by a solid sample which is measured by reflectance and other spectrum baseline affecting phenomena (Pasquini, 2003). Mathematical pretreatment

of spectra reduces noise or background information (smoothing techniques) and increases (multivariate) signals from the chemical information (differentiation) (Agelet & Hurburgh, 2010) so that they will better adhere to Beer's law (i.e., absorbance (spectral signals) are linearly correlated with analyte concentrations) (Rinnan et al., 2009). Hence, the goal of pretreatment is to eliminate or minimise variability unrelated to the property of interest so that pertinent changes can be more effectively modelled (Huang, Romero-Torres, & Moshgbar, 2010). It is expected that the data pretreatment process will improve the robustness and accuracy of subsequent quantitative or classification analyses; improve interpretability; detect and remove outliers and trends; and reduce the dimensionality of the data mining task (Lasch, 2012). The pretreatment of signals are generally determined by trial and error since they cannot be anticipated (García-Sánchez et al., 2017). Consequently, preliminary analyses are carried out to select the pretreatment that improves the signal/noise ratio of the spectra. The best spectral pre-processing is determined empirically based on the prediction performance for the calibration samples in cross-validation (Aernouts, Polshin, Lammertyn, & Saeys, 2011). The procedures for pre-processing the NIRS spectrum to obtain a good correlation between the spectral data and the concentration (García-Sánchez et al., 2017) are divided into two categories namely: the filtering methods and the model-based methods. The filtering methods, like calculating second derivatives or performing vector normalisation, simply transform spectra into a presumably better version of the same data, removing some undesired types of variation (Afseth & Kohler, 2012). Model-based methods, on the other hand, allow for the quantification and separation between the different types of physical and chemical variations in the spectra (Afseth & Kohler, 2012). The pre-processing methods used in the present study include: the extended mutiplicative scatter correction (EMSC) and the combination of Savitzky-Golay (SG) algorithm with EMSC.

Powders usually display scatterring (i.e., additive effect, multiplicative effect, and wavelength-dependent baseline variation), due to variation in particle size or shape or powder density, and this can be reduced by using appropriate sample presentation and/or the multiplicative scatter correction or EMSC for pre-processing the spectra (Mauer et al., 2009). It is critical to apply appropriate pretreatment to minimise such physical effects in order to be able to model the chemical effect more effectively (Huang et al., 2010).

EMSC is a model-based preprocessing technique that separates and provides quantitative estimates of the different types of chemical and physical variations in vibrational spectra, in

addition to the basic preprocessing of baseline correction and normalisation (Afseth & Kohler, 2012). It has been shown to be a reliable tool to correct for unwanted additive baseline effects, multiplicative scaling effects, and interference effects (Afseth & Kohler, 2012). The additive corrections are obtained by orthogonalising the spectra with respect to the directions representing irrelevant additive trends in the data (Skogholt, Liland, & Indahl, 2019). The multiplicative corrections are based on a chosen reference spectrum (i.e., the mean spectrum), and each original spectrum is appropriately scaled so that it can be expressed as a sum of the reference spectrum and a residual part representing the spectral information of interest (Afseth & Kohler, 2012; Skogholt et al., 2019). The purpose of the reference spectrum is to facilitate the estimation of multiplicative effects for transforming the measured spectra to a common scale (Skogholt et al., 2019). EMSC is particularly useful in minimising wavelength-dependent light scattering variation (Huang et al., 2010) and in removing spectral interferences of chemical origin (Bruun, Søndergaard, & Jacobsen, 2007). Thus, EMSC contributes to achieving better interpretability of the vibrational spectra, in addition to making the subsequent calibration models simpler and statistically more robust (Afseth & Kohler, 2012).

Derivation is a common way of removing the baseline features in a spectroscopic data and the derivatives may be calculated using the SG method (Afseth, Segtnan, & Wold, 2006). The SG routine can be used for smoothing/noise reduction, in addition to estimating the derivatives of the spectral data (Rinnan et al., 2009) to resolve overlapping signals, suppress unwanted spectral features and enhance signal properties (Zimmermann & Kohler, 2013). The SG smoothing/derivative method is carried out in one single step and it involves the idea of a convolute and of a convolution function (i.e., a vector of convolution integers and a normalisation factor) (Lasch, 2012). The convolute of a given spectra data point x_0 is obtained by computation of the dot product between the vector of convolution integers and a spectral segment of equal length with the midpoint being x_0, and dividing the result by a normalisation factor (Lasch, 2012). The SG method is based on fitting a simple polynomial (typically quadratic or quartic with lower degree preferred for noise reduction) to a small section of given size of the spectrum, and calculating the derivative of the polynomial in the centre point of the segment (Afseth et al., 2006). The polynomial order, and especially the window size, can strongly influence the properties of the derivated curve, and consequently the result of the multivariate analysis (Zimmermann & Kohler, 2013). The first derivative pre-processing

emphasises the steep edges of a peak and pronounces small features over a broad background (Mauer et al., 2009). Thus, the first derivative is expected to remove the offset features of the baseline while the second derivative is expected to handle the linear baseline features (Afseth et al., 2006).

The SG procedure is the most widely used smoothing algorithm in spectroscopy since it `attenuates high-frequency signals, such as noise, while at the same time tends to preserve important features of the chemical analyte signals, such as relative maxima, minima, height and width (Zimmermann & Kohler, 2013). However, the SG method is not able to remove multiplicative/scaling effects and as such the SG transformed data may be subsequently corrected by EMSC to enhance the analyte signals (Bruun et al., 2007). When the EMSC method is performed after the conversion of spectral data into a derivative form, it principally has the feature of removing the multiplicative effect since the broad baseline structures are effectively suppressed by the derivatives (Zimmermann & Kohler, 2013). It has been demonstrated that when EMSC is used alongside the SG procedure, the derivative treatment of data by the SG algorithm must precede the EMSC normalisation (Zimmermann & Kohler, 2013).

However, several hundreds or even thousands of variables/features are measured in an NIR spectrum and some of them may be redundant (i.e., irrelevant to the analysed problems or uncorrelated to the response) (Chen et al., 2018). Hence, there is a great need to do feature/variable selection in order to construct high quality models with minimal errors in prediction (Chen et al., 2018). The use of variable selection procedures before the modelling step have demonstrated their value, contributing to achieving good models, better and more interpretable than is possible by indiscriminate use of all the variables (Pasquini, 2003). The objective of variable selection is three-fold: to improve the prediction performance (i.e., robustness) of the predictors, to provide faster and more cost effective predictors, and to provide a better understanding of the underlying process that generated the data (Guyon & Elisseeff, 2003). According to Pasquini (2003), the selection of relevant variables from a large multivariate set is an area of pretreatment of NIR data that has undergone notable evolution. To determine whether some variables (wavelengths) could be removed to make the models more robust (fewer principal components or latent variables) and better performing (lower root mean squared error of cross validation) (Aernouts et al., 2011), two families of methods are usually considered: filter and wrapper methods. Filter methods (e.g., the variable

importance in projection) use an evaluation function that relies solely on the properties of the data without dependence on any particular algorithm (Talavera, 2005). These methods use the statistical properties of the data to filter out poorly informative features prior to the modelling stage and as such they are independent of model construction (Chen et al., 2018). On the other hand, the wrapper methods use the inductive learning algorithm to select subsets of features that are useful to build a good predictor (Guyon & Elisseeff, 2003). Although the wrapper methods are more computational demanding than the filter methods (Chen et al., 2018), they are widely recognised as a superior alternative in supervised learning problems, since by employing the inductive algorithm to evaluate alternatives, they have taken into account the particular biases of the algorithm (Talavera, 2005). For this study, only the wrapper methods such as the interval partial least-squares regression (iPLS) technique will be explored. The purpose of iPLS, proposed by Nørgaard et al. (2000), is to find one or a few intervals (i.e., windows of variables) which give better predictions than the predictions obtained when using the full spectrum, and to provide an overview to help in spectroscopic interpretation (Andersen & Bro, 2010). To implement the iPLS technique, the spectra are divided into a number of intervals of equal length and the model performances of the intervals are compared based mainly on their root mean squared error of cross-validation (RMSECV) but also on the number of components (those with minimum RMSECV value are chosen) and the correlation between measured and predicted values (Andersen & Bro, 2010).

2.3.3 Sampling strategies

The design of NIRS chemometric models begins with the selection of representative samples using the calibration sampling algorithms. Consequently, the models are developed on a representative portion of the data (training set) and validated on the remaining set of samples (test/validation set) (Stevens & Ramirez-Lopez, 2014). There is no fixed number or rule-of-thumb to determine the number of samples to be included in a calibration (Agelet & Hurburgh, 2010). However, between 20 and 200 samples are necessary to develop a multivariate calibration method but the greater the number of samples, the more representative and robust the calibrations will be (García-Sánchez et al., 2017). Calibrations of homogeneous mixtures (i.e., pharmaceutical powders) may require smaller calibration sets than agricultural samples of high compositional complexity and heterogeneity, such as whole grains or forages (Agelet & Hurburgh, 2010). The samples should have a normal distribution, cover the entire range of concentrations of the parameters that are to be estimated by NIRS, and should not

have areas where uncertainty is high (García-Sánchez et al., 2017). The labour-effective, inexpensive, ease and speed of NIR technique makes it feasible to screen the large number of samples necessary to ensure a representative training set (Shetty, Rinnan, & Gislum, 2012). The actual prediction performance of a calculated model should be evaluated on an independent test set that is representative of the future samples (Aernouts et al., 2011). There are several sampling strategies (Stevens & Ramirez-Lopez, 2014) but only the random selection (RS) sampling method and the Kennard-Stone (KS) algorithm technique (Kennard & Stone, 1969) will be used in this study.

The RS sampling method is a way of data splitting in which random division is applied without any clear selection criteria (Saptoro, Tadé, & Vuthaluru, 2012). Hence, when using the RS method, there is a risk that valuable objects of some classes are not selected in the training set, and that risk may be avoided by selecting 3/4 of the objects from each class separately, then putting them together as the training set (W. Wu et al., 1996). The KS algorithm technique is a type of data splitting in which all objects are considered as candidates for the training set and the selected candidates are chosen sequentially (Saptoro et al., 2012). The aim at each stage of the procedure is to select the objects so that they are unformly spaced over the object space (W. Wu et al., 1996). The technique takes the pair of samples with the largest Eucledian distance of x-vectors (predictors) and then it sequentially selects a sample to maximise the Eucledian distance between the x-vectors of already selected samples and the remaining samples; the process is repeated until the required number of samples is achieved (Saptoro et al., 2012). The KS algorithm technique allows to create a calibration set that has a flat distribution over the spectral space and the method used to compute the distance between points can be either the Euclidean distance or the Mahalanobis distance (Stevens & Ramirez-Lopez, 2014).

2.3.4 Qualitative spectral analysis

Qualitative analysis highlights patterns/features in the pre-processed spectral data and classify them into separate groups based on their identity or quality attribute using the supervised and non-supervised learning algorithms. Thus, it is divided into i) data structure analysis and ii) classification modelling to differentiate between sample groups and to characterise their spectra while detecting departures from the dataset. Qualitative analysis is performed using spectral library matching or pattern recognition algorithms such as the PCA to reduce the amount of data present in the pre-treated spectra in order to 1) facilitate the

classification and calibration steps, 2) handle multicollinearity and 3) to get a better overview of the data (Metrohm AG, 2014). Table 2.2 reported the various qualitative statistical techniques that have been applied in different levels of investigations using the NIRS. However, in this thesis, PCA is used as the first step of the exploratory analysis for data structure definition. PCA is a technique that allows data visualisation by the reduction of data dimensionality, retaining as much information in the original data as possible (Berrueta, Alonso-Salces, & Héberger, 2007). In brief, PCA is an unsupervised pattern recognition method as well as an exploratory tool (Cho & Kang, 2011) to reduce the number of highly correlated observed variables to a smaller number of uncorrelated principal components which account for most of the variance of the observed variables (Suhr, 2005). It is an unsupervised clustering or data reduction technique which is particularly useful in identifying how one sample is different from another and which variables contribute most to this difference, and whether those variables contribute in the same way (i.e. are correlated) or independently (i.e. uncorrelated) from one other (Wishart, 2007). PCA projects the spectral data to a new reduced dimensional space defined by axis called the principal components (PC) (Agelet & Hurburgh, 2010). PC are consecutively created following the directions of data variability in decending order while preserving their orthogonality constraints (Agelet & Hurburgh, 2010). PCA is not a particularly reliable approach for discrimination and classification since it is only able to identify gross variability and is not capable of distinguishing 'among groups' and 'within groups' variability (Barker & Rayens, 2003). Hence, the dimension reduction provided by PCA is guided only by total variability (Barker & Rayens, 2003).

The exploratory analysis of the spectral data is followed by classification modelling based on the pattern of their measured features. The purpose of classification is to study the global properties of a group of similar objects in order to build a mathematical model that will be used to assign future unlabelled samples to specific class fingerprints. The classification rules can be built using the supervised (e.g. linear discriminant analysis (LDA), partial least square – discriminant analysis (PLS-DA), soft independent modelling of class analogy (SIMCA), support vectors machine (SVM)) or the unsupervised (independent component analysis (ICA), k-nearest neighbour (k-NN), Mahalanobis distance classification (MDC)) methods. Roggo, Duponchel, and Huvenne (2003) compared the efficacies of the available qualitative statistical methods in classifying sugar beet through near infrared spectroscopy and they showed that SIMCA, PLS-DA, and the Procrustes discriminant analysis (PDA) had the highest classification

accuracies. In another study, Kong, Zhang, Liu, Nie, and He (2013) showed that PLS-DA and k-NN obtained over 80 % accuracy in classifying rice seed cultivar whereas SIMCA and SVM had 100 % classification accuracy in both the calibration and prediction set. According to Pasquini (2003), SIMCA is the most employed technique for qualitative analysis using NIRS. Hence, in this study, SIMCA (S. Wold & Sjostrom, 1977) and PLS-DA (S. Wold, Ruhe, Wold, & Dunn, 1984) are used for the classification modelling of the quality and safety of the milk products.

Table **2.2:** Qualitative and quantitative methods adopted for spectral data analysis

Investigation	**Qualitative**	**Quantitative**	**Reference**
Aged mushroom	PCA, PLS-DA	LS-SVM	Xu, Fu, et al. (2015)
Milk variances	PCA		Botros et al. (2013)
Infant formula & milk adulteration	PCA	PLSR, Poly-PLSR, ANN, LS-SVM	Balabin and Smirnov (2011)
Milk adulteration	PLS-DA	LS-SVM	Lu et al. (2009)
Milk adulteration	Factor analysis	PLSR	Mauer et al. (2009)
Quality loss in pasteurised skim milk	PCA, SIMCA	PLSR	Al-Qadiri et al. (2008)
Milk adulteration	PCA	PLSR, LS-SVM	Borin et al. (2006)

SIMCA was the first class modelling technique (i.e., sample assignations involving one or more predefined classes) used in chemometrics, and the central feature of this non-probabilistic distance-based modelling method is the application of PCA to the sample category studied, generally after within-class autoscaling (i.e., centring each variable and then dividing by standard deviation) or mean centring (i.e., subtracting the avarage from each variable) (Oliveri, Di Egidio, Woodcock, & Downey, 2011). The main steps of SIMCA analysis include: independent modelling of each group using PCA, calculation of the class distance (i.e., the geometric distance from the PC models), and calculation of the modelling power (i.e., the importance of the variables in modelling the class) and the discriminatory power (i.e., the variables that are important in discriminating between two classes) (Brereton, 2003). SIMCA models are defined by the range of the sample scores on a selected number of low-order PC (i.e., significant PC define the SIMCA inner space) and the PC not used to describe the model define the outer space considered as uninformative space (Oliveri et al., 2011). The number of PC for each class in the training set is determined by cross-validation (Berrueta et al., 2007). With SIMCA, two classes can overlap (and hence are ‘soft’), and there is no problem with an object belonging to both (or neither) class simultaneously (Brereton, 2003).

PLS-DA is a partial least square regression (PLSR) of a set *Y* of binary variables describing the categories of a categorical variables on a set *X* of predictor variables (Pérez-Enciso & Tenenhaus, 2003). It is a discriminant approach (i.e., sample assignations involve at least two predefined classes) that aims to find the variables and directions in the multivariate space which discriminate the established classes in the calibration set (Berrueta et al., 2007). The method provides a linear delimiter applying PLSR using binary class membership indices (e.g ., 0 and 1) for each class as the response variable (Oliveri, 2017). Hence, in PLS-DA, a 'dummy' *Y* matrix is constructed with zeros and ones (an observation had the value 1 for the class it belongs to and 0 for the rest) and the *X* matrix consists of the original (pre-processed) data (Berrueta et al., 2007). Matrices *X* and *Y* are decomposed in a product of other two matrices of scores and loadings (i.e., the loadings of the *X* block are calculated from the scores of the *Y* block, while the loadings of the *Y* block are determined from the scores of block *X*) ((Berrueta et al., 2007). The closer an element of a certain column in Y is to 1 and the elements of the other columns to 0, the more likely an object is a member of the particular class (Berrueta et al., 2007). Although PLSR was not inherently designed for problems of classification and discrimination, it performs well in that role and has heurestic support owing to the relationship between PLSR and canonical correlation analysis (CCA) and the relationship, in turn, between CCA and linear discriminant analysis (Barker & Rayens, 2003). PLSR performs well as a classification tool because it provides dimension reduction that is guided explicitly by among-groups variability in a discriminant application (Barker & Rayens, 2003).

2.3.5 Quantitative spectral analysis

Quantitative analysis establishes correlations between the pre-processed spectral data and the target parameters. Beer's law defines the concentration of the sample in terms of the path length, absorptivity and concentration; with the intensity of the absorbance peaks varying with concentration (Mauer et al., 2009). Hence, the quantitative analysis correlates the reference analytical data (i.e., the concentrations) with the calibration spectra (i.e., the absorbances) over a specified wavelength range using statistical techniques such as the partial least square regression (PLSR). The essence of the calibration is to train the NIR model to mimic the reference method.

PLSR is a method for relating two data matrices, *X* (i.e., the independent variable) and *Y* (i.e., the dependent variable), by a multivariate model, and for modelling their structure (S. Wold, Sjöström, & Eriksson, 2001). The principle of PLSR is to find the components in the input matrix

(*X*) that describe as much as possible the relevant variations in the input variables and at the same time have maximal correlation with the target value in *Y*, given less weight to the variations that are irrelevant or noisy (Berrueta et al., 2007). Consequently, PLSR models both *X* and *Y* simultaneously to find the latent variables in *X* that will predict the latent variables in *Y* while maximsing the covariance between matrices *X* and *Y* (Berrueta et al., 2007). PLSR has been shown to be a powerful technique for multivariate calibration when the data contain noise, large variable size (even if variable dimensionality is more than the sample size), highly collinear variables and missing observations, although it is strongly influenced by extreme samples/outliers (Shetty et al., 2012). Hence, PLSR derives its usefulness from its ability to analyse data with many, noisy, collinear, and even incomplete variables in both *X* and *Y* (S. Wold et al., 2001). PLSR is applicable even under conditions of very small sample sizes (Shetty et al., 2012). The various quantitative methods that have been successfully applied in related investigations are reported in Table 2.2 but only PLSR will be explored in this research project to quantify the milk components, the storage time and the levels of adulteration of the FMP samples formulated with different backbone fats. Balabin and Smirnov (2011) compared the applicability of different multivariate statistical algorithms on large industrial data sets and they found that the quantitative calibration methods such as the PLSR were capable of building robust models if they were used with correct data pre-processing or pre-treatment (e.g. data transformation and variable selection) technique for the spectral analysis. The authors stated that PLSR was the most important linear calibration method. An important feature of PLSR is that it takes into account errors in both the concentration estimates and the spectra (Brereton, 2003). However, to achieve robust regression models, a small but sufficient number of latent variables or factors should be included (i.e., using the explained y-variances as important indicators) (Afseth et al., 2006).

Chapter 3

Paper I

Efficacy of near infrared spectroscopy to segregate raw milk from individual cows between herds for product innovation and traceability

Abstract

Cows with specialized characteristics and requirements can be aggregated into different herds for targeted nutritional management and to facilitate on-farm segregation of raw milk for the production of high value niche dairy products, offering improved economic returns. Rapid methods for independent verification of product quality and origin are desirable to support validation and traceability of such products. This study examined the use of near infrared spectroscopy (NIRS) to segregate raw milk from individual cows of multiple breeds from different herds fed on the same or differing feeding regimes; and to correlate and evaluate the efficacy of the predictions for crude protein and the milk fatty acid (FA) phenotypes for each of the herds. Reference values and near infrared spectra were obtained from representative freeze-dried raw milk samples (n = 220) collected from 847 lactating cows of 3 breeds from the Lincoln University dairy farm in New Zealand. The feed sources (i.e., pasture or pasture with lucerne silage) significantly influenced the protein and the FA values and these differences were reflected in NIRS analyses. The partial least square regression models for crude protein determination showed excellent results whereas that for the most dominant FA were not appreciable. Maximum separation was obtained between the herds on the same feeding regime (mean specificity = 95.2 %) using the partial least square discriminant analysis and its overall performance in differentiating the objects was better than that of the soft independent modelling of class analogy. The multiclass analyses conducted in this study offer improvements to current approaches for evaluating and validating raw milk for the manufacture of specific dairy products, and for enhancing product traceability.

Keywords: milk segregation, near infrared spectroscopy, partial least square, soft independent modelling

3.1 Introduction

The composition of raw milk is influenced by breed, nutrition, health status and time of lactation of individual cows in the herds (Katz et al., 2016). Consequently, the composition of bulk milk (i.e. the sum of the milk contributed by all the animals that milked into the storage vats) can also vary. The recent volatility in the global dairy trade has prompted producers to reduce dependency on low value bulk commodities, to focus on premium segregated milk products that meet the consumer preferences. There is significant potential in segregating raw milk between herds to assist the manufacture of specific/differentiated/high value dairy products (Dooley et al., 2005), as well as providing quality assurance and traceability benefits. However, on-farm 'made-to-order' segregation of raw milk by herds is not yet common practice by farmers in New Zealand, despite these commercial and logistic advantages. Nevertheless, the key nutrients of fat and protein in cow milk vary tremendously between herds and they can be influenced through nutrition and feeding management (Heinrichs, Jones, & Bailey, 1997). In recent years, consumers have come to consider animal diet as a major criterion to estimate the quality of dairy products (Andueza, Agabriel, Constant, Lucas, & Martin, 2013), with those based on grass enjoying positive image and demand (Martin et al., 2004). Thus, to meet the growing demand of consumers who are prepared to pay premium prices for tailored products, grass-fed milk may be differentiated between herds as such segregation imbues value. In addition, tracing the herds of origin of raw milk may be valuable in the instances of fraud and contamination. There remains a need for analytical methods to independently verify milk quality prior to sale or distribution.

Several researchers have successfully characterized the nutritional quality of raw milk from different feeding systems based on their dominant FA profiles (Collomb, Bütikofer, Sieber, Jeangros, & Bosset, 2002; Hurtaud, Dutreuil, Coppa, Agabriel, & Martin, 2014). However, those characterizations involved cumbersome and expensive analytical reference methods such as gas chromatography (GC) which are generally less suitable for extensive routine authentication of dairy products (Coppa et al., 2012).

Multi-parametric instruments based on near infrared spectroscopy (NIRS) have been developed to provide precise, rapid, low-cost and non-destructive analytical solutions and they have been shown to have potential when combined with chemometric tools for characterizing cow milk according to feeding systems and origin (Coppa et al., 2012). However, those studies used bulk milk samples collected from different regions without any reference

to the variances between individual cows and their herds of origin. Consequently, there are no guarantees that the baseline quality of such samples was preserved right from milking until the commencement of the laboratory analysis. It is well known that the genetic background of any given herd could also affect the milk traits (i.e., within breed variation) (Marchitelli et al., 2013). So far, to the best of our knowledge, no studies have been undertaken to assess the potential of NIRS to differentiate the nutritional quality of cow milk from different herds on an ongoing basis at individual animal level. In our study, we applied NIRS to freeze-dried raw milk samples from individual dairy cows from mixed breeds, which were aggregated into different herds under differing feeding regimes, in a carefully controlled rearing environment in New Zealand and examined the classification and correlation with FA and crude protein contents, to assess the utility of the approach in differentiating milk quality between herds. Thus, in this paper, different chemometric methods such as principal component analysis (PCA), soft independent modelling of class analogy (SIMCA), partial least square discriminant analysis (PLS-DA) and partial least square regression (PLSR) were adopted to analyze and to make key decisions from the near infrared spectral datasets.

3.2 Material and methods

3.2.1 Sample collection, handling and preparation

A total of 50 mL sample of fresh raw milk was collected during the morning (0600 h) and afternoon (1430 h) milking processes from each of the 220 lactating cows from four herds (i.e., 56 in herd 1; 64 in herd 2; 49 in herd 3; and 51 in herd 4) on differing feeding regimes using inline milk meters (DeLaval International AB, Tumba, Sweden) as part of a 2-year biomonitoring study which commenced in September 2013 at Lincoln University Research Dairy Farm. The dairy cows (Friesian, Jersey and Friesian x Jersey) in herds 1 and 2 calved in late winter (August) whereas those in herds 3 and 4 calved in mid spring (October). The cows were mixed age (between 3 and 10 years old), in good health and in the middle of their lactation period. The mean liveweights for the cows in herds 1, 2, 3, and 4 were 507 ± 48.0 kg, 477 ± 55.0 kg, 499 ± 45.0 kg and 474 ± 50.0 kg, while their days in milk (DIM) were 273 ± 25.0 days, 259 ± 30.0 days, 265 ± 23.0 days and 263 ± 27.0 days respectively. The treatment grouping was equitably done to ensure that there was a balance within the herds for calving date, milk yield, days in milk and liveweight while maximizing the spread between herds. Herds 1 and 2 were fed mainly pasture with the addition of lucerne (*Medicago sativa*, alfalfa)

silage during the 2013-14 milking season; whereas herds 3 and 4 were fed only pasture during the 2014-15 season. The pasture was composed mainly of standard mixtures of perennial ryegrass (*Lolium perenne L*.) and white clover (*Trifolium repens*) base. All cows in the same herd were fed with the same diet throughout the milking season. The main difference between the herds was the style of their nutritional management which in principle was designed to meet the herd-specific requirements in order to optimize efficiency and milk quality. The raw milk samples, collected from individual cows, were tagged with their identities and thoroughly mixed to ensure uniformity. Then, they were frozen at -21^{o}C without any preservative before freeze drying (Cuddon Ltd New Zealand, Model E.D.5.3) at 4^{o}C in preparation for spectra measurements using the near infrared spectrophotometer. The freeze-dried milk samples were stored immediately at 4^{o}C until analysis.

3.2.2 Compositional analysis by primary methods

3.2.2.1 Milk crude protein determination

Crude protein content was measured in this study using the combustion (Dumas) method implemented using the Rapid Max N Exceed Elemental Analyser (Elementar GmbH, Germany). In brief, about 200 mg of the freeze-dried milk samples were weighed in duplicates into the stainless-steel crucibles and combusted at 900^{o}C in an oxygen atmosphere. The combustion process converted any elemental nitrogen into N_2 and NO_x; NO_x moieties were then reduced to N_2 while the other volatile combustion products were trapped and eliminated. Thereafter, the N_2 gases were passed through a thermal conductivity cell in order to measure the total nitrogen. The crude protein content of the sample was then calculated by applying the nitrogen-to-protein conversion factor of 6.38 and it was expressed in percent (or g per 100 g) of freeze-dried milk powder.

3.2.2.2 Milk fatty acid determination

The method employed for the determination of FA concentrations was based on two major phases, namely: one-step methylation and GC analysis as previously described (Ejeahalaka & On, 2019a). In brief, 100 µL of internal standard (C21:0 ester) was added into each of the two Kimax tubes containing 0.15 g of the freeze-dried milk samples. Then, 900 µL of heptane and 4.0 mL of 0.5 M NaOH/dried methanol were added into the tubes followed by the addition of

2.0 mL of heptane and 2.0 mL of distilled water. The tubes were then capped, vortexed and centrifuged to separate the heptane layer of the FA esters.

The GC analysis was carried out by injecting 1.0 µL of the extracted FA methyl esters (FAME) onto a Varian CP7420, tailor-made fused silica capillary column using the AOC-20i auto-sampler fitted to a Shimadzu GC-2010 gas chromatograph. The individual fatty acids in the sample were quantified in mg using the GC peak areas of the internal (C21:0) and external standards together with the mass of internal standard in the sample and the response factor of the individual fatty acids. Results were expressed in absolute terms in g of fatty acid per 100 g of freeze-dried milk samples by dividing the individual FA in mg by the mass of the sample in g.

3.2.3 Spectral measurements

Near infrared spectroscopic measurements were carried out on the freeze-dried milk samples using a FOSS NIRSystem (model DS 2500F) scanning spectrometer as previously described (Ejeahalaka & On, 2019a). In brief, 5 g of each of the 220 samples was placed in a 10 mL DS 2500 ring cup (cup type: 2004) and scanned at 0.5 nm intervals in the wavelength range of 850 to 2500 nm in triplicates. A total of 3300 absorbance values were captured per NIRS spectrum averaging 32 scans. The triplicate spectra acquired for each sample were averaged and then subjected to multivariate chemometric data analysis after undergoing the necessary pre-treatments.

3.2.4 Spectral pre-processing and chemometric analysis

Open source software, R project for statistical computing, version 3.5.1 (R Core Team, 2018), was used to process the spectral data and to develop models for exploring the resolution of the freeze- dried raw milk samples to herd level. It was also used to analyze the significant differences between the mean values of the reference data (i.e., for crude protein and the FA). The NIRS data were first subjected to PCA after mean-centring to evaluate potential relationships or groupings of readings obtained from the milk samples derived from the different animals and herds examined. PCA has been shown to be useful in identifying how one sample is different from another and which variables contribute most to this difference, and whether those variables contribute in the same way (i.e. are correlated) or independently (i.e. uncorrelated) from each other (Lavine & Workman, 2004; Wishart, 2007). Consequently,

the PCA loadings were plotted to identify the most sensitive wavelengths in the spectral data. To build the classification rules and to predict the reference (i.e., crude protein and FA) concentration values, the raw spectra data were pre-processed using Savitzky-Golay (SG) algorithm (Savitzky & Golay, 1964) (window size = 55 points, second-order polynomial fit) followed by the Extended Multiplicative Signal Correction (EMSC) method (Martens & Stark, 1991) to preserve higher moments, remove baseline effects and to suppress wavelength-dependent light-scattering variations. Zimmermann and Kohler (2013) demonstrated that the use of SG with EMSC for pre-processing results in simpler and often better models, since the SG differentiation effectively suppresses the broad underlying baselines, while the EMSC principally has the feature of removing the multiplicative effect. The authors further showed that EMSC generally performed better on SG differentiated spectroscopic data than the multiplicative signal correction (MSC) because of the lack of higher terms in the MSC model as opposed to the linear and quadratic terms in the EMSC model. To proceed with the analyses, the pre-processed spectra data were split randomly in the ratio of 4:1 into training sets (i.e., 42 in herd 1, 48 in herd 2, 37 in herd 3 and 38 in herd 4) to build the calibration models, and test sets (i.e., 14 in herd 1, 16 in herd 2, 12 in herd 3, and 13 in herd 4) to measure the predictive ability of the models on unseen samples. The leave-one-out cross-validation method was used for internal validation of the training data in order to choose the models that will perform best on the test sets.

3.2.4.1 Quantitative prediction of raw milk phenotypes by herds

Partial least square regression is used for relating and modelling the structure of two data matrices, X and Y, with many noisy, collinear, and even incomplete variables (S. Wold et al., 1984). This supervised multivariate technique mathematically correlates the spectral intensities at different wavelengths (called the absorbance values) as the independent X variables to the reference sample composition (e.g. crude protein or fatty acid contents) as the dependent Y variable. The power of PLSR lies in its ability to derive latent variables (LV) or factors (i.e., linear combinations of the X variables) that correlate to as great an extent as possible with the Y variable as well as having maximal predictive powder on Y (Meilgaard, Carr, & Civille, 1999; H. Wold, 1966). Two approaches can be followed in carrying out a PLSR analysis namely: PLS1 (used when only one Y variable is predicted) and PLS2 (used when several Y variables are predicted). In this study, PLS1 approach was used for modelling the crude protein contents whereas PLS2 was used for modelling the selected five of the most dominant FA i.e.,

palmitic acid (C16:0), stearic acid (C18:0), oleic acid (C18:1 c9), rumenic acid (C18:2 c9,t11), and α-linolenic acid (C18:3 c9,c12,c15) and the FA groups namely: saturated (SFA) and the unsaturated fatty acids (UFA) in the freeze-dried raw milk samples. The models were built using the full spectra range (850 – 2500 nm) and the windows of spectral wavelengths that contained the most relevant information as determined by interval-PLS (Nørgaard et al., 2000). The predictive performances of the regression models were evaluated using the root mean squared error-observations standard deviation ratio (RSR) as previously described (Ejeahalaka & On, 2019b). Although lower values of the root mean squared error of prediction (RMSEP) indicate higher model predictive accuracy, RSR provides a delimiter of what is considered a low RMSEP based on the standard deviation of the measured values (Moriasi et al., 2007). RSR standardizes RMSEP using the observations standard deviation (Moriasi et al., 2007) and as such it is the reciprocal of the residual prediction deviation (RPD) of the model. RSR varies from the optimal value of 0, which indicates zero RMSEP and therefore perfect model prediction, to a large positive value (Moriasi et al., 2007) However, as RPD > 2 is desired for a good calibration (Karoui et al., 2006), so also is RSR < 0.5.

3.2.4.2 Raw milk near infrared segregation by herds

PLS-DA and SIMCA were employed to segregate the raw milk samples into their herds of origin. PLS-DA is a supervised discriminant classification approach requiring at least two classes to be defined, (Barker & Rayens, 2003) and it is essentially based on the PLS2 algorithm that searches for LV with a maximum covariance with the Y variables (H. Wold, 1966). However, the Y-block in PLS-DA is a qualitative variable as it describes which objects are in the classes of interest (H. Wold, 1966). On the other hand, SIMCA is specifically designed to study and describe one single class at a time and it checks for compatibility of unknown samples with the class being modelled (Oliveri & Downey, 2012; S. Wold & Sjostrom, 1977). SIMCA is a supervised classification (or pattern recognition) method that helps to demarcate the boundaries of predefined classes and to assign future objects to those classes. In this work, the PLS-DA and the SIMCA models were built using both the full spectral range (850 – 2500 nm), and the subset of the original dataset covering the sensitive bands selected from the PCA loadings. The classification performances of the models were evaluated using the sensitivity, specificity, correct classification rate, precision, and the Matthews correlation coefficient as previously defined in the literature (Ballabio & Todeschini, 2009; Oliveri, 2017).

3.3 Results and discussion

3.3.1 Characteristics of NIRS profiles

Figure 3.1 shows the mean NIRS values of the freeze-dried raw milk samples from individual cows from each of the four herds in the near infrared region from 850 – 2500 nm. Spectral profiles of milk from each of the herds displayed similar absorption characteristics, with some features typical for NIRS analyses of oven-dried goat milk (Núñez-Sánchez et al., 2016) and full cream milk powder (Frankhuizen, 2001). However, differences between milk from herd types receiving differing nutrition were seen, with generally higher absorbance values observed in milk from cows fed on pasture alone. This finding correlates with those of Mouazen, Dridi, Rouissi, De Baerdemaeker, and Ramon (2009) who reported higher absorbance values for pasture feeding as compared to soya-bean feeding in their study of the properties of ewe milk. These authors described the milk as having higher density and lesser water content.

Based on previous research (Frankhuizen, 2001), six bands that are most useful for analysing the major constituents of milk powders are distinguishable in Figure 3.1 as indicated at about 1649, 1722, 1758, 1934, 2218 and 2308 nm in each of the spectra. The PCA loadings plot (Figure 3.3) also identified the range from 1580 to 2305 nm as the optimal wavelengths and therefore was included in our models. It is likely that the constituents contributing to absorption at these bands correspond to: water, protein, and lactose (1650 nm); protein and fat (1734 nm); fat and protein (1759 nm); water (1940 nm); fat, protein and lactose (2230 nm); and fat and protein (2310 nm) as previously reported (Frankhuizen, 2001). Also, the three typical water absorption bands that characterize the milk spectrum are distinguishable at about 950 nm, 1445 nm and 1934 nm.

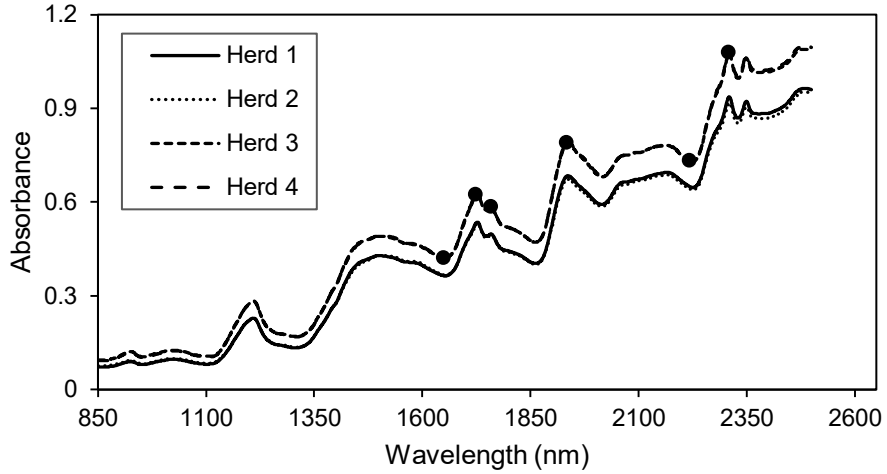


Figure **3.1:** Average near infrared spectra of the freeze-dried raw milk for each of the herds (● = most useful selected bands for milk analytics).

3.3.2 NIRS analysis with PCA

The PCA performed on the 3300 independent spectral variables (Figure 3.2) of the freeze-dried milk powder supported the mean profile analysis in displaying substantive separation of samples taken from herds fed with different sources. The score plots showed an overlap between the objects of herds 1 and 2 and also between those of herds 3 and 4, thus requiring the use of multivariate classification methods (i.e. SIMCA and PLS-DA) for their possible differentiation.

The loadings variable plot showing the wavelengths that had the largest effect on PC 1 and PC 2 is presented in Figure 3.3. The identified wavelengths include: 921, 1210, 1390, 1580, 1722, 1945, 2100, 2305 and 2345 nm. However, only the wavelength range 1580 – 2305 nm was considered optimal for building the chemometric models as they yielded the most stable predictions.

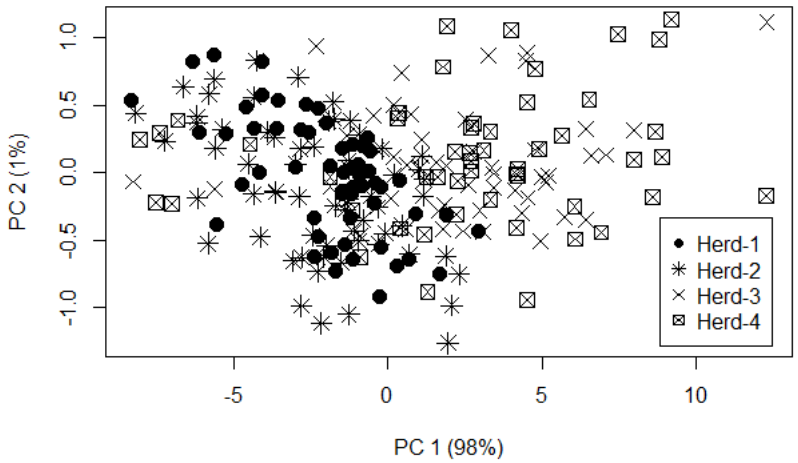


Figure **3.2:** PCA Score plot of the near infrared spectra for the 4 herds of the dried raw milk samples.

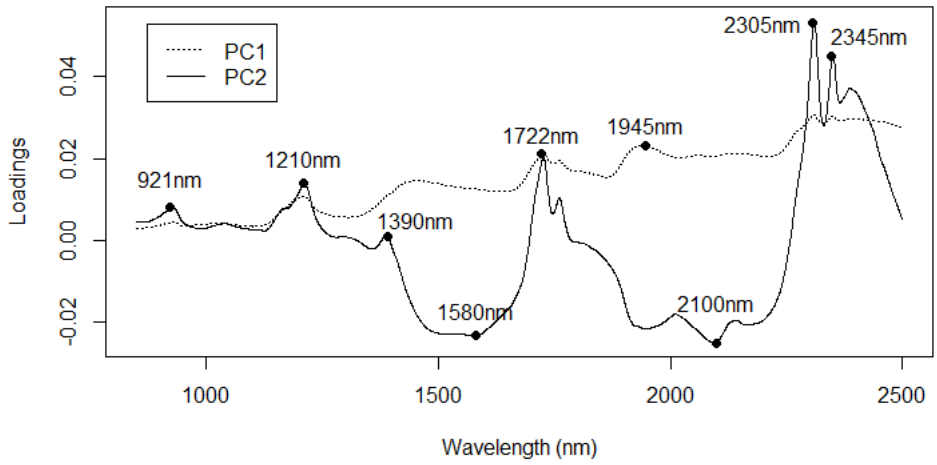


Figure **3.3:** Loadings variable plot of the first two principal components, PC 1 and PC 2, showing the optimal wavelengths

3.3.3 Reference data analysis and the PLSR modelling of milk phenotypes

It can be seen from Table 3.1 that the total UFA for herds 2 and 3 were statistically higher than those for herds 1 and 2 whereas the total SFA remained the same for all the herds.

Table **3.1:** Fitting statistics of the NIRS prediction models for the dominant fatty acids of herd-classified raw milk using development data sets expressed on per milk basis as g/100 g of milk.

			Prediction performance metrics[2]			
	Descriptive stats[1]		SG + EMSC Full		SG + EMSC iPLS	
Fatty Acids	Mean	CV	R2p	RSR	R2p	RSR
C16:0						
Herd 1	12.47[a]	16.17	0.64	0.58	0.65	0.57
Herd 2	13.32[a]	17.64	0.68	0.73	0.70	0.68
Herd 3	12.96[a]	22.90	0.67	0.58	0.71	0.55
Herd 4	12.71[a]	25.13	0.80	0.45	0.86	0.44
C18:0						
Herd 1	3.20[bc]	20.44	0.03	1.00	0.11	0.91
Herd 2	2.93[c]	19.68	0.16	1.03	0.18	1.12
Herd 3	3.52[ab]	23.35	0.40	0.88	0.52	0.72
Herd 4	3.69[a]	28.99	0.55	0.89	0.62	0.80
C18:1c9						
Herd 1	4.72[b]	15.46	0.32	0.85	0.45	0.72
Herd 2	4.44[b]	14.96	0.28	1.21	0.32	1.19
Herd 3	6.99[a]	15.83	0.72	0.61	0.72	0.67
Herd 4	7.03[a]	23.45	0.32	0.81	0.63	0.62
C18:2c9t11						
Herd 1	0.43[b]	29.12	0.05	0.96	0.19	0.88
Herd 2	0.29[c]	40.02	0.35	1.00	0.53	0.76
Herd 3	0.59[a]	41.78	0.16	0.92	0.39	0.76
Herd 4	0.59[a]	42.60	0.10	0.94	0.28	0.82
C18:3c9c12c15						
Herd 1	0.27[b]	17.35	0.08	1.00	0.13	0.99
Herd 2	0.28[b]	16.72	0.34	0.81	0.36	0.80
Herd 3	0.32[a]	28.23	0.28	0.92	0.47	0.82
Herd 4	0.33[a]	27.08	0.22	0.86	0.37	0.82
SFA						
Herd 1	24.93[a]	13.58	0.51	0.76	0.54	0.71
Herd 2	25.37[a]	15.17	0.73	0.77	0.74	0.79
Herd 3	25.95[a]	19.94	0.88	0.41	0.89	0.41
Herd 4	25.94[a]	22.95	0.88	0.50	0.91	0.44
UFA						
Herd 1	9.13[b]	11.92	0.22	1.00	0.50	0.70
Herd 2	8.82[b]	50.07	0.03	0.99	0.07	0.98
Herd 3	12.44[a]	10.99	0.84	0.50	0.87	0.40
Herd 4	12.52[a]	17.75	0.13	1.04	0.54	0.67

[a-c] Same superscript letters within a row are not significantly different at $p \leq 0.05$.

[1]Mean and coefficient of variation estimated for Herd 1 (n = 56), Herd 2 (n = 64), Herd 3 (n = 49) and Herd 4 (n = 51) with a total of 220 samples.

[2]SG + EMSC = Savitzky-Golay plus extended multiplicative signal correction pre-processing methods performed on full spectral range, 850 – 2500nm, of the calibration sets (n = 166) and using interval-PLS. R^2p = coefficient of determination for prediction (n = 54). RSR = RMSE-observation standard deviation ratio calculated as 1/RPD (ratio of prediction to deviation). Optimal RSR ≤ 0.5.

The inclusion of lucerne in the diet resulted in a statistically significant decrease in oleic acid (C18:1c9), rumenic acid (C18:2c9t11) and α-linolenic acid (C18:3c9c12c15). This result concurred with the findings of Dierking, Kallenbach, and Roberts (2010) and Rugoho et al.

(2014). Dierking et al. (2010) found from their study that grasses had higher amounts of α-linolenic acid compared with lucerne. Similarly, Rugoho et al. (2014) reported that the milk produced from cows consuming high-quality pasture contained high levels of α-linolenic acid and rumenic acid. Overall, the coefficients of variation of the FA concentrations were consistent with that of Fleming et al. (2017) and the values found for herds 3 and 4 were higher than those for herds 1 and 2.

The crude protein contents for the 4 herds of the freeze-dried raw milk samples are shown in Table 3.2. It can be seen from the table that the average values were higher for herds 1 and 2 than for herds 3 and 4. The recorded differences in mean values of the crude protein contents can be attributed to the effects of the inclusion of lucerne (a legume high in nitrogen) in the feeding regimes of herds 1 and 2. Woodward, Waghorn, Attwood, and Li (2010) showed in their study that a change from ryegrass to lucerne reduced milkfat and increased milk protein concentrations. The PLSR models for crude protein determination (Table 3.2) using the iPLS variable selection showed excellent results (RSR ≤ 0.41).

Table **3.2:** Fitting statistics of the NIRS prediction for crude protein contents (%) of herd-classified raw milk

			Prediction performance metrics[2]			
	Descriptive statistics[1]		SG + EMSC Full		SG + EMSC iPLS	
Crude Protein	Mean	Range	R2p	RSR	R2p	RSR
Herd 1	26.16[a] ± 1.15	24.09 - 28.01	0.85	0.53	0.90	0.41
Herd 2	25.59[a] ± 2.10	23.15 – 30.15	0.94	0.25	0.98	0.21
Herd 3	23.55[a] ± 3.33	20.05 – 30.56	0.92	0.25	0.93	0.23
Herd 4	23.70[a] ± 3.21	19.11 – 31.73	1.00	0.11	1.00	0.08

[a-c] Same superscript letters within a column are not significantly different at p ≤ 0.05.
[1]Mean ± standard deviation estimated for Herd 1 (n = 56), Herd 2 (n = 64), Herd 3 (n = 49) and Herd 4 (n = 51) using Dumas Method.
[2]SG + EMSC = Savitzky-Golay plus extended multiplicative signal correction pre-processing methods performed on full spectral range, 850 – 2500nm, of the calibration sets (n = 166) and using interval-PLS. R^2p = coefficient of determination for prediction (n = 54). RSR = RMSE-observation standard deviation ratio calculated as 1/RPD (ratio of prediction to deviation). Optimal RSR ≤ 0.5.

For FA determination (Table 3.1), the RSR values ranged from 0.40 to 1.21. In general, only 8 objects had RSR values ≤ 0.50 that was desired for good predictions. To provide more details, we observed that the PLSR models developed with the full spectral range (850 – 2500 nm) yielded 4 good predictions whereas those built with iPLS variable selection gave 4 optimal predictions. Thus, the selection of sensitive wavelength intervals for building the calibration models did not improve the overall prediction outcome. Among the individual FA, 16:0 had the lowest RSR across all the herds, with herd 4 having the best prediction (RSR = 0.44). The

minimum RSR recorded for the five C18 fatty acids predicted was 0.61 which was above the cut-off of 0.50 desired for good calibrations. Thus, the prediction accuracy of the PLSR models for the selected FA was inadequate. Nevertheless, there was a slight improvement in the global prediction performance across the herds for the FA groups. The RSR for UFA ranged from 0.40 to 0.98 whereas that for SFA ranged from 0.41 to 0.79 across the herds. Several researchers obtained similar loss of prediction accuracy for FA and they attributed it to the effects of the differences in breeds of individual cows on the sample matrices. Tsenkova, Atanassova, Itoh, Ozaki, and Toyoda (2000) obtained poorer results for near infrared determination of fat contents when the samples (in liquid form) from individual cows of multi-breeds were combined and predicted collectively than when they were simulated individually. These authors attributed the discrepancies in accuracy between the predictions to the differences in chemical composition and physical structure of the milk from each cow.

3.3.4 Multiclass chemometric differentiation of raw milk by herds

Table 3.3 and Table 3.4 show the respective model performance metrics for PLS-DA and SIMCA for multiple classification of the NIRS spectra of the freeze-dried raw milk samples into herds. Three different treatments of the calibration sets are presented. In the first treatment, no pre-processing was applied and the full spectra range (850 – 2500 nm, 3300 wavelengths) was used for the calibrations. In the second and third treatments, the same pre-processing (SG + EMSC) was applied but using the full spectra range (850 – 2500 nm) and the selected sensitive wavelength range (1580 – 2305 nm, 1451 wavelengths) for analyzing the major constituents of milk powder respectively.

Table **3.3:** Multi-class PLSDA classification of freeze-dried raw milk for 4 herds using different spectral pre-processing techniques

Model with no pretreatment[1]										
	Test sets assigned into classes				Model performance metrics[4]					
True Class	Herd 1	Herd 2	Herd 3	Herd 4	LV	SENS	SPEC	CCR	P	MCC
Herd 1	10/14	15/16	12/12	12/13	11	71.4	95.1	89.1	0.83	0.70
Herd 2	11/14	14/16	12/12	13/13	11	87.5	92.3	90.9	0.82	0.78
Herd 3	14/14	16/16	7/12	10/13	15	58.3	93.0	85.5	0.70	0.55
Herd 4	14/14	16/16	9/12	9/13	11	69.2	92.9	87.3	0.75	0.64
Model with pretreatment[2]										
Herd 1	12/14	14/16	12/12	13/13	12	85.7	95.1	92.7	0.86	0.81
Herd 2	12/14	14/16	12/12	13/13	8	87.5	94.9	92.7	0.88	0.82
Herd 3	14/14	16/16	7/12	10/13	12	58.3	93.0	85.5	0.70	0.55
Herd 4	14/14	16/16	11/12	8/13	13	61.5	97.6	89.1	0.89	0.68
Model with variable selection[3]										
Herd 1	11/14	14/16	12/12	13/13	9	78.6	95.1	90.9	0.85	0.76
Herd 2	12/14	15/16	11/12	13/13	5	93.8	92.3	92.7	0.83	0.83
Herd 3	14/14	16/16	6/12	8/13	14	50.0	88.4	80.0	0.55	0.40
Herd 4	14/14	16/16	10/12	7/13	5	53.8	95.2	85.5	0.78	0.56

[1]Full spectrum (850 – 2500 nm) models calibrated (n = 165) with no pre-processing; test sets n = 55.

[2]Full spectrum models pre-processed with Savitzky-Golay algorithm in addition to the extended multiplicative signal correction method.

[3]Varible selection models (1580 – 2305 nm) pre-processed with Savitzky-Golay algorithm in addition to the extended multiplicative signal correction method.

[4]LV = latent variables; SENS = sensitivity (%); SPEC = specificity (%); CCR = correct classification rate; P = precision; and MCC = Matthews correlation coefficient for evaluating the model efficiency.

Table **3.4:** Multi-class SIMCA classification of freeze-dried raw milk for 4 herds using different spectral pre-processing techniques

Model with no pretreatment[1]										
	Test sets assigned into classes				Model performance metrics[4]					
True Class	Herd 1	Herd 2	Herd 3	Herd 4	PC	SENS	SPEC	CCR	P	MCC
Herd 1	12/14	13/16	12/12	13/13	10	85.7	92.7	90.9	0.80	0.77
Herd 2	3/14	13/16	12/12	13/13	13	81.2	71.8	74.5	0.54	0.49
Herd 3	13/14	14/16	9/12	5/13	8	75.0	74.4	74.5	0.45	0.42
Herd 4	14/14	16/16	2/12	10/13	13	76.9	76.2	76.4	0.50	0.47
Model with pretreatment[2]										
Herd 1	13/14	6/16	11/12	13/13	5	92.9	73.2	78.2	0.54	0.58
Herd 2	4/14	13/16	12/12	12/13	8	81.2	71.8	74.5	0.54	0.49
Herd 3	14/14	15/16	11/12	1/13	5	91.7	69.8	74.5	0.46	0.51
Herd 4	10/14	15/16	1/12	12/13	5	92.3	61.9	69.1	0.43	0.46
Model with variable selection[3]										
Herd 1	13/14	8/16	12/12	12/13	4	92.9	78.0	81.8	0.59	0.63
Herd 2	3/14	15/16	12/12	13/13	11	93.8	71.8	78.2	0.58	0.60
Herd 3	14/14	16/16	11/12	2/13	4	91.7	74.4	78.2	0.50	0.56
Herd 4	14/14	16/16	0/12	12/13	8	92.3	71.4	76.4	0.50	0.55

[1]Full spectrum (850 – 2500 nm) models calibrated (n = 165) with no pre-processing; test sets n = 55.

[2]Full spectrum models pre-processed with Savitzky-Golay algorithm in addition to the extended multiplicative signal correction method.

[3]Varible selection models (1580 – 2305 nm) pre-processed with Savitzky-Golay algorithm in addition to the extended multiplicative signal correction method.

[4]LV = latent variables; SENS = sensitivity (%); SPEC = specificity (%); CCR = correct classification rate; P = precision; and MCC = Matthews correlation coefficient for evaluating the model efficiency.

As seen in the tables, the main diagonal elements represent fractions of true positives whereas the off-diagonal elements indicate fractions of false positives for each of the herds. For the first treatment, the mean values recorded across the herds for SENS, SPEC, CCR, P and MCC for the PLS-DA model were 71.6 %, 93.3 %, 88.2 %, 0.78 and 0.67 respectively, whereas those computed for the SIMCA model were 79.7 %, 78.8 %, 79.1 %, 0.57 and 0.54 respectively. For the second treatment, the mean values recorded across the herds for SENS, SPEC, CCR, P and MCC for the PLS-DA model were 73.3 %, 95.2 %, 90.0 %, 0.83 and 0.72 respectively, whereas those computed for the SIMCA model were 89.5 %, 69.2 %, 74.1 %, 0.49 and 0.51 respectively. Finally, for the third treatment, the mean values recorded across the herds for SENS, SPEC, CCR, P and MCC for the PLS-DA model were 69.1 %, 92.8 %, 87.3 %, 0.75 and 0.64 respectively, whereas those computed for the SIMCA model were 92.7 %, 73.9 %, 78.7 %, 0.54 and 0.59 respectively. Thus, the PLS-DA model had higher specificity and Matthews Correlation Coefficient, correct classification rate, and precision (i.e. implying smaller standard deviation and coefficient of variation). The SIMCA model only had higher mean sensitivity than the PLS-DA model and as such performed better in assigning objects to their native

classes/herds. However, the SIMCA model, with lower specificity, presented higher false positives while differentiating the objects from the herds that received the same diet in the same milking year. The implication of higher specificity for the PLS-DA model is that it provided better differentiation between the herds on the same feeding regime which were shown to be overlapped in Figures 3.1 and 3.2 (i.e. herds 1 and 2, and herds 3 and 4) with fewer false positives/misclassifications. This result has further echoed the findings of Dooley et al. (2005) that raw milk varies between herds making it possible to segregate it on-farm for the manufacture of specific dairy products. However, freeze-drying the samples for several days for *in situ* measurements may pose a problem of using NIRS as a quick method for on-farm segregation of raw milk. Our methodology was undertaken under small-scale laboratory conditions involving freeze-drying using available resources and it appears that raw liquid milk may alternatively be used in a compatible NIRS liquid analyser to achieve similar on-farm segregation efficiency. Andueza et al. (2013) showed in their study that they were no significant differences when fresh and freeze-dried cheeses were used to compare the ability of two NIRS methods to distinguish between pasture and preserved-forage cheeses. It has long been established ((De la Roza-Delgado et al., 2017) that handheld NIRS instruments can successfully be used *in situ* to estimate the changes in individual raw liquid milk composition. Furthermore, as the Matthews correlation is a contingency matrix method for calculating the Pearson product-moment correlation coefficient (Powers, 2011), it will invariably take the same range of values as follows as adopted from Chan (2003): at least 0.8 for very strong, 0.6 up to 0.8 for moderately strong, 0.3 to 0.5 for fair, and less than 0.3 for poor linear relationship. Thus, the performance of the PLS-DA model can be described as reasonably good while that of SIMCA was only fair. Overall, the pre-processed PLS-DA calibration using the full spectra range (treatment 2) provided the best model for this study. The other pre-processing techniques, for example treatment 3, could not better the outcome of treatment 2 using PLS-DA since they yielded lower CCR, P, and MCC values. In addition, the calibration performed using the raw spectra without pre-processing (i.e., treatment 3) did not improve the results.

3.4 Conclusion

The average near infrared spectra and the PCA established that there was a difference in the freeze-dried milk powder samples between herds on different feeding regimes, but it was not clear whether such differences exist between those on the same diet. This study has demonstrated that although SIMCA models achieved high sensitivity, they could not reliably

differentiate the near infrared spectra of the milk powder samples from different herds on the same feeding regimes where the PCA model had failed to describe the classes because of their superimposed objects. That implies that SIMCA models excelled only where PCA showed clear separation between the objects. However, the PLS-DA models were reasonably effective in differentiating between the herds on the same diets and their overall performance was better than that of SIMCA, thus highlighting the possibility of segregating raw milk on-farm between herds for the manufacture of specific dairy products and for enhancing product traceability. These results support value for NIRS for characterising milk and efficacy in validation of feeding regimes, which are useful for independent quality assurance. It has also been shown that the inclusion of lucerne in the diets significantly impacted on the quality of the milk powder. Furthermore, PLSR models were developed and they provided excellent predictions for crude protein contents but not for the selected fatty acid profiles. The poor accuracy of the fatty acid determinations resulted from the collective prediction of milk powder samples, with differing matrices, from individual cows of multiple breeds. However, a seldom used, but meaningful performance metric called “RMSE-observations standard deviation ratio” was successfully used in this study for the evaluation of the PLSR models. Our results indicate useful, independent and cost-effective quality analytics of milk powder to discriminate at herd level can be achieved that may be beneficial to manufacturers of high-value dairy products. Further study is required using the raw liquid milk samples for wider adoption and to improve the methodology described herein.

Chapter 4

Paper II

Chemometric studies of the effects of milk fat replacement with different proportions of vegetable oils in the formulation of fat-filled milk powders: implications for quality assurance

Abstract

Bovine milk lipids can be replaced with cheaper indigenous vegetable oils to produce milk alternatives with healthier saturated/unsaturated fat balance for those in areas where milk supply is poor or even absent. A wide range of vegetable oils can be used, but their impacts when blended with skimmed milk powder to formulate filled milk powder (FMP) are still unknown. We investigate the baseline variances in 12 types of FMP produced onsite with 3 proportions (10%, 20% and 30%) of 4 different vegetable oils (i.e., coconut, palm, soya-bean and sunflower) using fatty acid- and near infrared spectra profiles. Chemometric analyses revealed 8 significant overlapping clusters of FMP types but 100% classification efficiency was achieved. Sunflower oil, and particularly soya-bean FMP types had statistically the lowest indices of atherogenicity and thrombogenicity. This appears to be the first chemometrics study of FMP; the spectral analytical models used may be effective for product monitoring.

keywords: vegetable oils, fat-filled milk powder, near infrared spectroscopy, partial least square discriminant analysis, soft independent modelling

4.1 Introduction

Filled milk powder (FMP) is a recombined product formulated to resemble whole/full cream milk powder with the substitution of milkfat by a locally available, cost-efficient vegetable oil (Early, 1998). FMP is a cheaper milk alternative that enables millions of consumers to access dairy nutrition in areas including many African nations where milk supply is very low or even absent, with over 890,000 tonnes estimated to be produced globally in 2015 (European Commission, 2017). Indeed, since bovine milk fat is composed mainly of saturated fatty acids (SFA) which are associated with poor health outcomes (Keys et al., 1986), the production of FMP also provides an excellent opportunity to enhance the nutritional value of reconstituted milk (Jensen & Nielsen, 1982). However, the composition of milk fat is far more complex than most of the fats used for the substitution and as such using some vegetable oils for FMP may result in nutritionally inferior product (Jensen & Nielsen, 1982). In addition, the use of vegetable oils may also result in unwanted changes in both organoleptic properties and/or nutritive value due to the composition of the fatty acids (FA) in some of the fats (Jensen & Nielsen, 1982).

In the standard for blending skimmed milk powder (SMP) with vegetable fat to formulate FMP adopted in 2006 (last amended in 2014), the Codex Alimentarius Commission of the Food and Agriculture Organisation and World Health Organisation specified that FMP should contain a total fat and milk protein of not less than 26 % (w/w) and 34 % (w/w) respectively (Codex Alimentarius Commission, 2006). However, different oils possess differing ratios of saturated, mono-unsaturated and poly-unsaturated fats (Early, 1998). For instance, coconut and palm kernel oils (called the lauric oils), which contain high levels of medium- and long-chain SFA, are frequently used for FMP due to their high oxidative stability, good mouth feel, low melting point and bland taste (Ibrahim, 2011). Palm oil is also commonly used in FMP (Jensen & Nielsen, 1982; Vignolles, Jeantet, Lopez, & Schuck, 2007) for similar reasons coupled with the fact that it has a balanced fatty acid composition in which the level of SFA is almost equal to that of unsaturated FA (Siew, 2011). However, the concern about the elevation of serum cholesterol level by saturated fats and oils has led to attempts to formulate FMP with vegetable oils with a high content of unsaturated fatty acids (UFA). Jeon, Roberts, and Senecal (1992) investigated the possibility of using partially hydrogenated oils of soya-bean, sunflower, canola and cottonseed whereas Modler, Rippen, and Stine (1970) showed that lightly hydrogenated soya-bean oil can successfully be used to formulate FMP with reasonable

oxidative stability and yet containing significant quantities of UFA, including the essential FA. The choice of vegetable oil is thus crucial in formulating FMP; unfortunately, the literature on FMP has so far been very limited. We are unaware of studies to compare the variances in different formulations of FMP with respect to their FA profiles, or to report methods for their routine classification.

Multi-parametric instruments (i.e., those allowing simultaneous analysis of multiple parameters) involving near infrared spectroscopy (NIRS) have been shown (Botros et al., 2013) to have potential in characterising the spectral variances of commercial SMP and nonfat dry milk powders when coupled with chemometrics. The benefits of NIRS compared to other technologies include its ready availability, low cost, high throughput, and rapid analytical measurements (Botros et al., 2013). NIRS appears to be ideal for objective characterization of FMP products for quality assurance, yet we have not found evidence of its application in this regard.

In this study, we describe what we believe to be the first use of NIRS analyses to characterize FMP samples formulated using 3 different proportions of 4 types of vegetable oils as the backbone fats. The baseline variances of the fatty acid profiles of the formulated FMP were also determined to aid understanding of the PCA, and the impacts on nutritive potential.

4.2 Material and methods

4.2.1 Raw materials collection

This study was conducted using three (i.e., palm oil, soya-bean oil and sunflower oil) of the four major vegetable oils in the world; and a lauric oil (i.e., coconut oil) which has a very different FA composition from other commodity oils (Gunstone, 2011). Palm oil (PO) was purchased from Pure Nature (New Zealand) whereas coconut oil (CO), soya-bean oil (SBO), sunflower oil (SFO) and SMP were purchased from a local supermarket chain in Christchurch, New Zealand.

4.2.2 Filled milk powder manufacture

A total of 240 FMP samples were formulated by first dispersing and mixing 25 g of SMP in 250 mL of deionised water at 20 ^{o}C using a magnetic stir bar while avoiding the incorporation of air. The four vegetable oils (i.e., CO, PO, SBO and SFO) used for the formulation were pre-warmed to 15 ^{o}C above their respective melting points, and different proportions (i.e., 10 %,

20 % and 30%) of each one of them were independently added to the reconstituted SMP. Then, the mixture was homogenised (Polytron, PT 3100) for 5 minutes at 6600 rpm while maintaining a constant temperature of 70 ^{o}C. The resulting emulsions were rapidly frozen (Moffat Friginox, Model RC30-15A) at – 32 ^{o}C for 2 hours without adding any emulsifier before they were stored in frozen state at – 20 ^{o}C for at least 2 days. Thereafter, the deep-frozen emulsions were subjected to freeze drying (Cuddon Ltd New Zealand, Model E.D.5.3) for a period of 10 days at 4 ^{o}C in preparation for NIRS measurements. The dried FMP samples were thoroughly mixed and ground to uniform particle sizes using a combination of a small burr coffee grinder and a porcelain spatula before they were immediately stored at 4^{o}C. Twenty samples were formulated for each of the 3 different oil proportions of the 4 FMP types, resulting in 240 samples for analysis. Our methodology was undertaken under small-scale laboratory conditions involving freeze drying and grinding, and likely to differ from those used by large-scale commercial manufacturers. Nonetheless, each of the oil supplements have been or are used to produce FMP available for sale.

4.2.3 Fatty acids extraction and detection in filled milk powders

The method employed for the determination of FA concentrations was based on two major phases: (a) one-step methylation and (b) gas chromatography (GC) analysis. The one-step methylation of the FA was done by weighing 0.15 g of the freeze-dried FMP samples directly into Kimax tubes in duplicates. Then, 100 µL of internal standard (C21:0 ester) was added into each of the tubes, followed by the addition of 900 µL of heptane and 4.0 mL of 0.5 M NaOH/dried methanol. The tubes, with their caps screwed, were carefully vortexed and incubated in water-bath at 50 °C for 15 minutes. Then, the tubes were vortexed again after cooling at room temperature, and that was followed by the addition of 2.0 mL of heptane and 2.0 mL of distilled water. The tubes were then capped, vortexed and centrifuged at 1500 g for 5 min in order to separate the heptane layer of the FA esters. The residual water in the heptane layer was subsequently removed using small amounts of anhydrous sodium sulphate and a subsample of the recovered extract called FA methyl esters (FAME) was stored in a vial at - 20 °C until GC analysis.

The GC analysis was carried out by injecting 1.0 µL of the extracted FAME onto a Varian CP7420, tailor-made fused silica capillary column with a length of 100 m, internal diameter of 0.25 mm, and film thickness of 0.20 µm, using the AOC-20i auto-sampler fitted to a Shimadzu

GC-2010 gas chromatograph. The chromatograph was operated under the following analytical conditions: carrier gas, helium with linear velocity of 16.7 cm/s; oven temperature, 45 °C to 250 °C (after 4 min at the initial temperature of 45 °C, the oven was ramped at the rate of 13 °C/min to 175 °C and was held for 27 min before ramping up again at the rate of 4 °C/min. It was held at 215 °C for 35 min and was baked-off at 250 °C for 5 min. The equilibration time between runs was 5 min); inlet, split injection with a ratio of 1:60; flame ionisation detector temperature, 250 °C; needle wash, heptane (2 pre-wash, 4 post-wash). The external standards used for identification and quantification of the individual fatty acid methyl esters were: ME 61, ME 93, ME 100, BR 2, BR 3, GLC 463, GLC 411 and CLA c9t11 (Larodan Fine Chemicals, Sweden). The FA concentrations obtained from the assay were expressed both as g of fatty acid per 100 g of total fatty acids and mg of fatty acid per g of freeze-dried milk samples; only the latter converted to g per 100 g of dried milk was used for analysing quality and baseline variances in this study.

4.2.4 Lipid nutritional implications characterisation indices

The nutritional implications of the lipid fractions of the FMP samples were assessed using the indices of atherogenicity and thrombogenicity proposed by Ulbricht and Southgate (1991). The two indices highlighted the principal fatty acids that are culpable for increasing cholesterolaemia (i.e., lauric (C12:0), myristic (C14:0), and palmitic (C16:0) acids) and the formation of thrombi (i.e., C14:0, C16:0, and stearic (C18:0) acid) respectively. According to Ulbricht and Southgate (1991), the index of atherogenicity (IA) was defined as:

$$IA = \frac{C12{:}0 + (4 \times C14{:}0) + C16{:}0}{\sum(n-6\ PUFA) + \sum(n-3\ PUFA) + \sum(MUFA)} \quad (4.1)$$

whereas the index of thrombogenicity (IT) was defined as:

$$IT = \frac{C14{:}0 + C16{:}0 + C18{:}0}{0.5\sum(MUFA) + 0.5\sum(n-6\ PUFA) + 3\sum(n-3\ PUFA) + \left(\frac{n-3\ PUFA}{n-6\ PUFA}\right)} \quad (4.2)$$

These indices will help to determine whether the replacement of milk fat with each type of vegetable oil enhances the milk nutritional quality by altering positively the FA composition

and as such was justified. The scores obtained for IA and IT would be highest for the most atherogenic and thrombogenic FMP respectively. Lower scores are indicative of weaker association of FMP lipids with the incidence of coronary heart disease and vice versa (Ulbricht & Southgate, 1991).

The FA relative ratios such as palmitoleic (POA, C16:1 c9): palmitic acid (PA, C16:0), oleic (OA, C18:1 c9): PA, linoleic (LA, C18:2 c9, c12): linolenic acid (ALA, C18:3 c9, c12, c15), and MUFA: SFA that are of interest for human nutrition and for quality authentication were also estimated.

4.2.5 Near infrared spectral measurements

The freeze-dried FMP samples were first exposed to laboratory conditions (temperature (25° C) and relative humidity (50 %)) for 2 hours in airtight containers and allowed to equilibrate. Then, 5 g of each of the milk samples was placed in a 10-mL DS 2500 ring cup (cup type: 2004) and scanned using a FOSS NIRSystem (model DS 2500F) spectrometer at a resolution of 0.5 nm, wavelength accuracy of < 0.05 nm, analysis time of < 1 minute and a detector array which was composed of both silicon (850 – 1100 nm) and lead sulfide (1100 – 2500 nm). A total of 240 unique samples (i.e., 20 from each of the 3 different oil proportions of the 4 FMP types) were scanned by the same operator in two consecutive days of sampling without shutting down the machine and each spectrum acquired comprised of 3300 absorbance values recorded from 850 to 2499.5 nm averaging 32 scans. Spectra were collected consecutively in triplicates from each sample using the same ring cup without refilling at each scan and their average was exported and subjected to chemometric analysis (described below). The triplicate spectra acquired for each sample were used to ascertain the device repeatability. Hence, the coefficients of variations were computed for the absorbance values, wavelength by wavelength, across the 3 replicate spectra.

4.2.6 Statistical analysis

The Tukey's Honestly Significance Difference test (Minitab 18 Statistical Software, 2017) was used to compare the differences between the mean values of the lipid indices, and the sum of the fatty acid compositions of different FMP types. The raw spectral data was transferred into the R software version 3.5.1 (R Core Team, 2018) for multivariate statistical analysis. In brief, Principal component analysis (PCA) was performed as the first step of the multivariate

analysis to identify groupings within the spectral data. Subsequently, classification modelling was used to evaluate the integrity and accuracy of unknown samples assigned to specific groups (Oliveri & Downey, 2012). Two approaches were used in this study namely: soft independent modelling of class analogy (SIMCA) and partial least square discriminant analysis (PLS-DA). The raw and pre-processed spectral data were used in each case for the classification modelling. For modelling with raw data, the entire spectra range (850 – 2500 nm, 3300 wavelengths) was used whereas only the sensitive spectra range was used for building the models with the pre-processed data. The extended multiplicative signal correction (EMSC) technique (Martens & Stark, 1991) was used for the pre-processing. The quality/applicability of the models obtained was evaluated by using external validation in which the raw data was randomly divided into a training set (to build the model) and a testing set (to test the model) in the ratio of 4:1. For the SIMCA model, the optimal number of principal components was determined through internal validation using the leave-one out cross-validation method whereas for PLS-DA, the optimal number of latent variables was determined by internal validation using random cross-validation (number of segments = 5).

The performances of the classification models were evaluated based on their sensitivity (SENS), specificity (SPEC) and efficiency. SENS (also known as the true positive or recognition rate) is defined as the fraction of the samples belonging to the modelled class which is correctly accepted or recognized by the model; whereas SPEC (also known as true negative or rejection rate) is defined as that fraction of samples not belonging to the modelled class that is correctly rejected by the model (Oliveri & Downey, 2012). The efficiency (E) of the model is a summarising parameter which is given as the geometric mean of the values of sensitivity and specificity (Oliveri & Downey, 2012), defined as follows:

$$E = \sqrt{Sensitivity \times Specificity} \tag{4.3}$$

4.3 Results and discussion

4.3.1 FMP variance by FA analysis

The baseline variances of FMP types containing 10 %, 20 % and 30% proportions of vegetable oils (CO, PO, SBO and SFO) were characterised by FA analysis and the results obtained are

shown in Table 4.1. As can be seen from the table, the FA of the FMP increased significantly with increasing proportions of the incorporated vegetable oils. However, the increment was disproportionate, and the total FA were statistically the same for CO and PO, PO and SBO, and for SBO and SFO for each of the incorporation levels (10 %, 20 % and 30 %). That implies that the FMP samples may not be adequately discriminated based only on their total FA.

According to Lee, Noh, Bae, and Kim (1998), vegetable oils have characteristic FA compositions that are useful for evaluating product quality and authenticity. Consequently, it was observed (Table 4.1) that when CO was the incorporated vegetable oil, the FMP samples had about 91.7 % SFA with C12:0 as the most dominant, contributing about 51.4 % of those FA. In addition, significant amounts of the *Capra* FA (i.e., C6:0, C8:0, and C10:0) were uniquely detected in those samples. However, POA and ALA were not detected at all in their FA composition. Thus, in formulating FMP, CO had the effects of raising the contents of C8:0, C10:0, C12:0, C14:0, and LA while lowering those of C6:0, C18:0, PA, POA, OA and ALA as compared to that of freeze-dried raw milk (data not shown). When PO, a *beta-prime* tending oil (i.e. containing a greater variety of fatty acids than *beta* tending oils, and more thermodynamically stable and forming smaller crystals), was incorporated in FMP to replace the milk fat, the samples were found to contain nearly equal amounts of SFA (55.8 %) and unsaturated FA (UFA) (44.2 %) with PA and OA being the most dominant FA respectively. In addition, the contents of C20:0, PA, OA and LA were raised while those of C6:0, C8:0, C10:0, C12:0, C14:0, C18:0, POA and ALA were lowered as compared to their values in freeze-dried raw milk. On the other hand, the inclusion of SBO and SFO, which are *beta* tending oils, in FMP resulted in 84.0 % and 89.5 % UFA respectively in the milk powder samples, with OA and LA as the most dominant FA. Nevertheless, the SFO FMP type contained higher values of OA and LA but with a comparatively insignificant amounts of ALA. However, the replacement of milk fat with either SBO or SFO had the same effects of raising the contents of C20:0, C22:0, OA, LA and ALA while reducing those of C6:0, C8:0, C10:0, C12:0, C14:0, C18:0, PA and POA as compared to their values in freeze-dried raw milk. These results showed that for each of the milk fat replacements, the incorporated vegetable oil had the common effect of raising the LA contents. In addition, none of the formulated FMP samples presented a value for conjugated linoleic acid.

Overall, the CO FMP type had the highest medium chain FA and the lowest OA and LA contents. It also had the most SFA. The inclusion of PO resulted in FMP with the highest PA

and C18:0 contents whereas the inclusion of SBO yielded FMP with the highest ALA contents. The SFO FMP type had the highest OA and LA contents and it was the most unsaturated in the present study. In addition, it had the highest MUFA and PUFA, owing to their high OA and LA contents.

In general, these results suggest that although vegetable oils differed in FA compositions, they lacked chemical markers that can be used for their immediate identification in the FMP samples. Moreover, different vegetable oils are normally blended together in the formulation of FMP to improve physicochemical and nutritional properties, and that may introduce unknown variances which pose a challenge in characterising the milk powder samples by FA analysis.

Table **4.1:** Fatty acid[1] composition (expressed as g/100 g of milk) of four types of filled milk powders[2] (FMP) formulated with different proportions (10%, 20% and 30%) of four types of vegetable oils

	10% veg. oil inclusion				20% veg. oil inclusion				30% veg. oil inclusion			
FA[3]	CO	PO	SBO	SFO	CO	PO	SBO	SFO	CO	PO	SBO	SFO
SFA												
C6:0	0.2	-	-	-	0.3	-	-	-	0.3	-	-	-
C8:0	3.1	-	-	-	4.4	-	-	-	4.9	-	-	-
C10:0	2.3	-	-	-	3.3	-	-	-	3.6	-	-	-
C12:0	22.6	0.1	-	-	30.9	0.2	-	-	34.2	0.2	-	-
C14:0	9.6	0.4	0.1	0.1	12.8	0.6	0.1	0.1	14.2	0.7	0.1	0.1
C16:0	4.9	22.4	5.0	3.4	6.4	33.2	8.0	5.0	7.0	39.4	9.6	5.6
C18:0	1.5	2.4	1.8	1.8	1.9	3.5	2.8	2.5	2.1	4.1	3.4	2.9
C20:0	0.0	0.2	0.2	0.1	0.1	0.2	0.3	0.2	0.1	0.3	0.3	0.2
C22:0	-	0.0	0.2	0.4	-	0.0	0.3	0.5	-	0.0	0.4	0.6
C24:0	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.2	0.0	0.0	0.2	0.2
MUFA												
POA	-	0.1	0.0	0.1	-	0.1	0.1	0.1	-	0.1	0.1	0.1
OA	3.4	15.0	10.2	18.1	4.1	22.1	16.3	24.8	4.5	26.0	19.6	28.2
GA	0.0	0.0	0.1	0.1	0.0	0.1	0.2	0.1	0.0	0.1	0.2	0.2
PUFA												
LA	1.0	5.2	25.2	31.5	1.1	7.6	41.1	47.3	1.2	9.0	49.5	53.9
ALA	-	0.1	2.3	0.1	-	0.1	3.8	0.2	-	0.2	4.6	0.2
Groups												
SFA	44.3	25.6	7.3	5.9	60.0	37.7	11.7	8.5	66.4	44.7	14.0	9.6
MUFA	3.5	15.1	10.4	18.2	4.1	22.2	16.6	25.0	4.5	26.2	19.9	28.4
PUFA	1.0	5.3	27.5	31.6	1.1	7.7	44.9	47.5	1.2	9.1	54.1	54.1
UFA	4.4	20.4	37.9	49.9	5.2	29.9	61.4	72.5	5.7	35.3	73.9	82.6
SUM	48.7[a]	46.0[a]	45.2[a]	55.8[a]	65.3[a]	67.6[a]	73.1[ab]	81.0[b]	72.1[a]	80.0[ab]	87.9[bc]	92.1[c]

[a,b,c]Means for the sum of the fatty acids with different superscripts differ significantly ($P < 0.05$).

[1]Values represent the mean of triplicate gas chromatography measurements.

[2]CO, PO, SBO and SFO represents the coconut oil, palm oil, soya-bean oil and sunflower oil FMP types.

[3]FA = fatty acid; POA = palmitoleic acid; OA = oleic acid; GA = Gondoic acid; LA = linoleic acid; ALA = alpha linolenic acid; SFA = saturated FA; MUFA = monounsaturated FA; PUFA = polyunsaturated FA; UFA = unsaturated FA; SUM = total of all the fatty acids in each of the FMP types.

4.3.2 FMP near infrared spectra characterisation

The coefficients of variations of the absorbance values, wavelength by wavelength across the triplicate spectra were < 3.5 %, indicating that the samples were considerably homogeneous, and the measurements were of acceptable repeatability. The representative raw spectra of the FMP samples produced with different proportions (10 %, 20 % and 30 %) of vegetable oils (CO, PO, SBO and SFO) are shown in Figure 4.1. It was observed that the average spectrum (i.e. mean of the absorbance values across the wavelengths for all the samples) for each of the FMP types followed a similar pattern, with minor differences and as such may be difficult to differentiate only by visual inspection. The maximum absorbance values recorded for CO, PO, SBO, and SFO were 1.004, 1.167, 1.766 and 1.868 respectively. The order of the absorption intensities for the FMP types was as follows:

SFO > SBO > PO > CO.

This variance may be related to their LA content since it was the only fatty acid that followed such a sequence in the present study. Thus, the higher the unsaturation of the vegetable oil included in the FMP, the higher the absorption intensity of the milk sample. It was also observed that increasing the proportions of the incorporated vegetable oils (except 30 % CO and 30 % PO) resulted in an increment in absorption intensity.

A close inspection revealed that some of the spectra of the FMP produced with CO overlapped with those produced with PO and that probably suggests that both vegetable oils share certain characteristics in common. A similar trend was observed in the spectra for FMP produced with SBO and those with SFO. In general, the FMP spectra were dominated by three strong absorption peaks which increased in intensity and broadness with increasing unsaturation in lipid and they were located at around 1210 nm, 1726 nm and 2313 nm. Each of those peaks was accompanied by a shoulder located at around 1170 nm, 1765 nm and 2350 nm respectively. The bands at 2313 nm and 2350 nm may be due to combinations involving methylene C–H stretching (Holman & Edmondson, 1956) whereas those at 1726 nm and 1765 nm may arise from the first overtone of C–H stretching vibration of methyl, methylene, and ethenyl groups (Westad, Schmidt, & Kermit, 2008). The weaker absorption bands around 1170 nm and 1210 nm are likely due to the second overtones of C–H stretching vibrations (Holman

& Edmondson, 1956; Westad et al., 2008; Woodcock, Downey, & O'Donnell, 2008). The FMP samples containing vegetable oils with *cis* double bonds were found to have moderate combination absorption band at around 2150 nm and weak second overtones at around 1170 nm possibly indicating their degree of unsaturation in lipid (Holman & Edmondson, 1956; Westad et al., 2008). The band at around 1395 nm appeared like a small lipid shoulder in all the spectra and have been attributed to the combination of C–H stretching and other vibrational modes (Westad et al., 2008). A relatively definable absorption band arising from third overtone of C–H stretching of fat was further observed at around 928 nm and that concurred with the wavelength assigned by (Šašić & Ozaki, 2001a) for raw milk samples in the shortwave near infrared region. The absorption bands due to –NH groups in protein were found near 2065 nm and 2180 nm in the FMP spectra whereas that due to –OH groups of water were found around 1450 nm and 1940 nm. The 2065 nm band was due to the N–H stretching combinations while the 2180 nm band was associated with the N–H bend second overtone and C=O stretch/N–H in-plane bending/C–N stretch combination bands (Workman & Weyer, 2007).

Overall, the spectral profiles of the FMP types look similar in every respect except for the visible differences in their absorption intensities and the presence of a shoulder around 1170 nm, 1385 nm and 2150 nm relating to the degree of unsaturation of the incorporated vegetable oils. However, CO and PO FMP types seem to have spectral profiles similar to that of freeze-dried raw milk powder as shown in Figure 4.1.

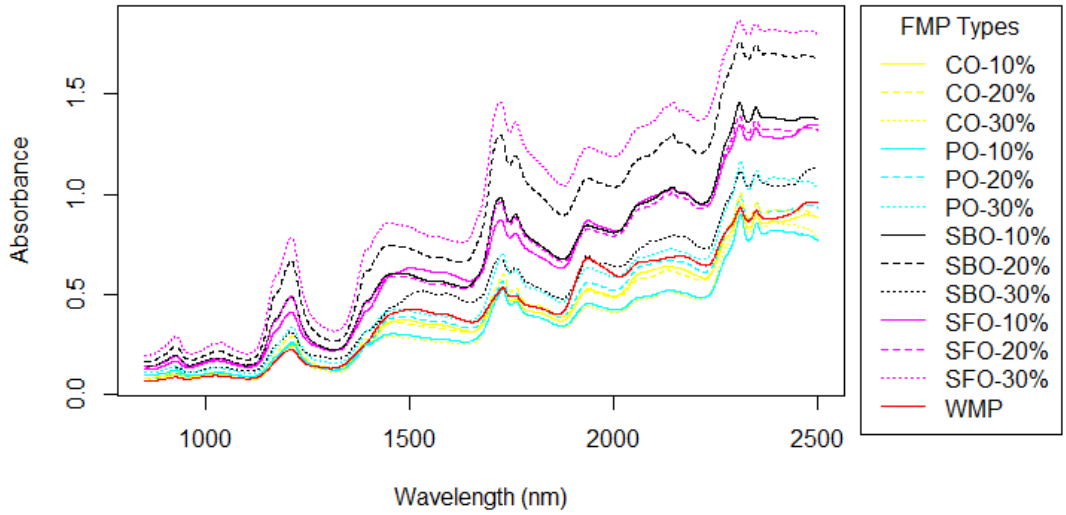


Figure **4.1:** Representative near infrared spectra of the 4 filled milk powder (FMP) types formulated with 3 proportions (10%, 20% and 30%) of 4 different vegetable oils; in addition to the spectra of freeze-dried raw milk powder (WMP). CO, PO, SBO and SFO represent the coconut oil, palm oil, soya-bean oil and sunflower oil FMP types respectively.

4.3.3 PCA filled milk powder characteristics

The trends observed in the spectral profiles of FMP were explored by PCA using the full, mean-centred NIR spectra (3300 variables) of the 240 unique milk powder samples without any pre-processing. The score plot for the PCA models is shown in Figure 4.2. The first two principal components (PC 1) and (PC 2) were retained for the PCA and they accounted for 99 % and 1 % of the total spectral variance respectively. Consequently, the dimensionality of the samples was reduced to two principal components as that were enough to retain 100 % of the original variance in the 3300 spectral variables. On the contrary, PC 1 and PC 2 only contributed a total of 99 % to the whole variation when the spectra data was standardised (i.e., mean-centred and scaled) prior to PCA analysis (data not shown). As can be seen from the score plot (Figure 4.2), there was a trend in which the FMP samples increased in scores along the co-ordinate of PC 2 with increasing proportions of CO and PO as well as along PC 1 with increasing proportions of SBO and SFO. The FMP samples containing CO and PO were clustered closely together according to treatments (i.e., oil proportions) and appeared separated from those having SBO and SFO (except for 10 % SBO) along PC 1, thus revealing an important structure in the spectral data. It is important to note that this clustering pattern coincided with the results of the total FA in Table 4.1. These observations implied that there was a treatment

effect and the CO and PO FMP types had similar characteristics but differed from those with SBO and SFO inclusions. In particular, PC 1, the most dominant component, separated the FMP samples into two groups according to their fat saturation, thus describing the direction of maximum variance, and the true underlying distributional structure of the spectral data. Hence, the FMP types were ordered according to their fat saturations in the horizontal directions as follows: CO > PO > SBO > SFO.

Further examination revealed that all samples containing the same type and proportion of vegetable oils were positioned close to one another and as such are similar except for those with 30 % SBO and 30 % SFO which probably had high variability owing to their dispersed lipid matrices. The fact that the samples containing CO overlapped with those with PO of the same proportion signified that both vegetable oils had some similarities in their compositions. Similar resolved clustering and overlap was noted for the FMP samples with 20 % SBO and 20 % SFO inclusions. Thus, no distinct separations were observed between most of the individual FMP types (especially those with 20 % and 30 % oil inclusions) and as such chemometric classification methods were needed to be able to characterise each of the sample groups under investigation.

The loadings variable plot showing the contributions of the original variables to PC 1 and PC 2 is presented in Figure 4.2 (lower diagram). The wavelengths corresponding to the highest loadings weights include: 1216, 1330, 1396, 1589, 1711, 1725, 1940, 2096 and 2304 nm. The wavelengths 1216, 1396, 1711, 1725 and 2304 nm were most likely correlated with C-H bond absorption in fats, whereas those of 1589 and 2096 nm may be related to the absorption of N-H groups in protein, as previously discussed in section 4.3.2. Thus, most of the differential features between spectra were contributed by the differing fat contents in our samples that originated from the various oils used to supplement our SMP in the FMP blends. However, the wavelength range 1216 to 1589 nm was chosen as optimal for analysis as it yielded the most stable and optimized predictions.

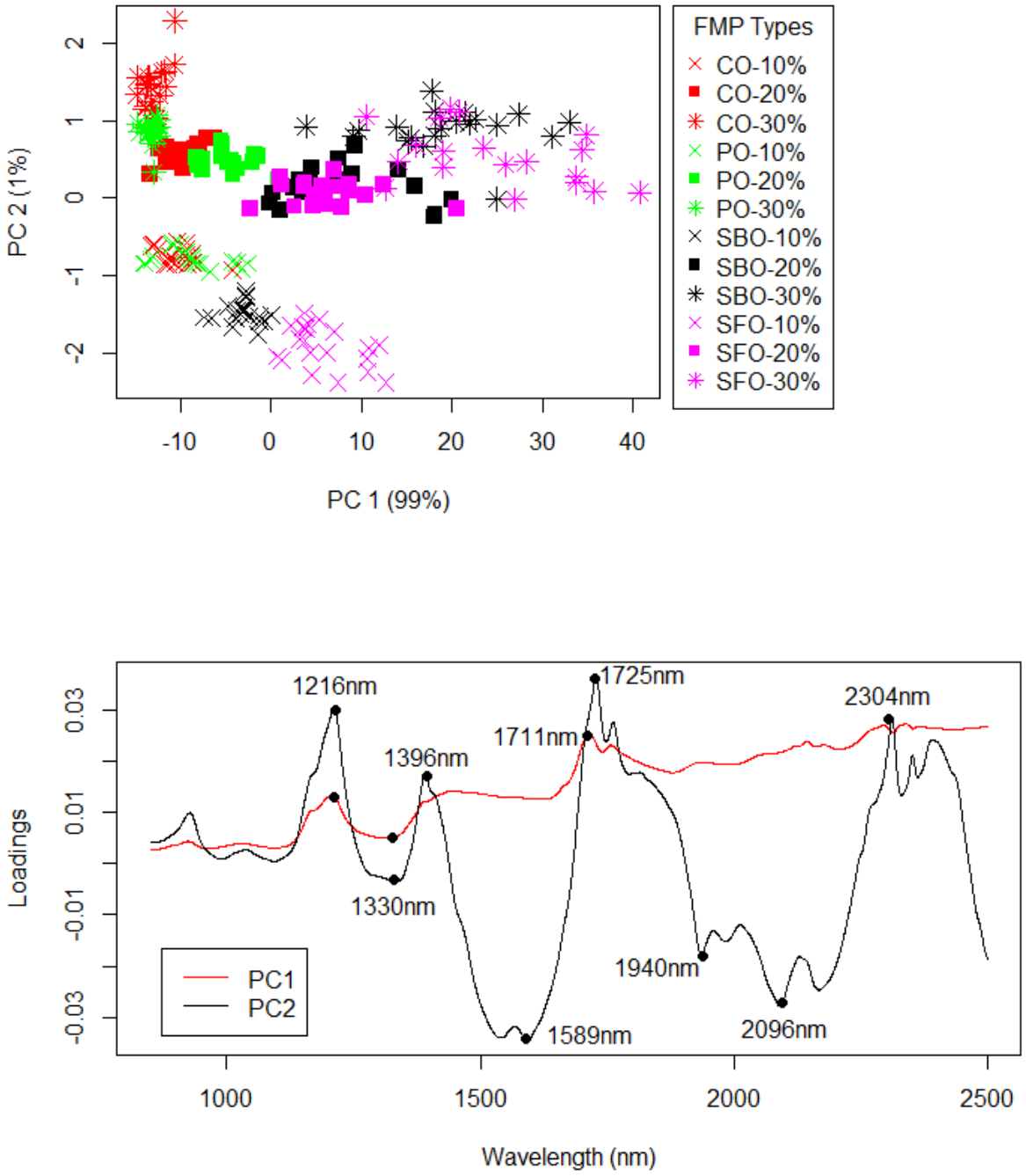


Figure **4.2:** Upper diagram shows the PCA score plot of the near infrared spectra of the filled milk powder (FMP) types formulated with 3 proportions (10%, 20% and 30%) of 4 different vegetable oils. CO, PO, SBO and SFO represent the coconut oil, palm oil, soya-bean oil and sunflower oil FMP types respectively. Lower diagram illustrates the sensitive wavelengths in the loadings variable plot of the first two principal components, PC 1 and PC 2.

4.3.4 Multiclass classification by supervised pattern recognition techniques

Table 4.2 and Table 4.3 summarised the recognition (sensitivity) and rejection (selectivity) abilities of SIMCA and PLS-DA respectively in classifying the test sets of the FMP types containing 20 % and 30 % proportions of the four different vegetable oils (i.e., CO, PO, SBO and SFO). The bold-faced diagonal elements represent the sensitivities whereas the off-diagonal elements indicate the specificity values. Two different treatments of the calibration sets are presented. In the first treatment, the raw spectra data (850 – 2500 nm, 3300 wavelengths) was used for building the calibration models without any pre-processing. In the second treatment, only the sensitive spectra range (1216 to 1589 nm, 747 wavelengths) was used for building the calibration models and in addition, an EMSC pre-processing technique was applied. As can be seen from the tables, the second treatment gave optimised results and will therefore only be discussed here.

For the SIMCA models, a maximum of four PC was needed to develop the classification rules and the results (Table 4.2) showed that the FMP types containing CO and PO (either 20 % or 30 %) were clearly distinguishable from the other classes without any overlap as they recorded maximum (100.0 %) SENS and SPEC values with 100.0 % efficiencies. However, the model for SBO FMP type performed poorly in classifying the test samples containing 30 % of SFO. Similarly, the efficiency of the model for SFO FMP type was not appreciable in classifying the test sets containing SBO and SFO irrespective of the proportion of the oil inclusion. This poor result could be attributed to the similarities in sample matrices of SBO and SFO. In general, the SIMCA models recorded mean SENS, SPEC and efficiency values of about 90.0 %, 95.0 % and 92.0 % across the entire classes respectively. This is a significant result considering that the PCA (Figure 4.2) conducted on the same spectral data yielded some discernible clusters with limited separations between the respective classes.

On the other hand, the PLS-DA models required a maximum of thirteen LV that best described the spectral data to build the classification rules. As can be seen in Table 4.3, the models correctly identified all the test sets that belonged to the modelled classes irrespective of treatments (i.e., vegetable oil types and proportions) resulting in zero false negatives and recognition rates (sensitivity values) of 100.0 %. Also, the models correctly rejected the test sets that did not belong to the modelled classes irrespective of treatments resulting in zero false positives and rejection rates (selectivity values) of 100.0 %. Thus, the PLS-DA models

recorded mean SENS, SPEC and efficiency values of 100.0%, 100.0 % and 100.0 % respectively across the entire classes. PLS-DA, therefore, achieved better efficiency values (reaching 100.0 %) than the SIMCA models (92.0 %) in classifying the FMP types. These results supported the assertion by Pomerantsev and Rodionova (2018) that PLS-DA favoured the separation of classes with the same major components (e.g. SMP) and different impurities (e.g. vegetable oils). Also, according to S. Wold, Eriksson, Trygg, and Kettaneh (2004), discriminant analysis works better where there is homogeneity and similarity between the classes as it was the case for our FMP samples. However, Azcarate, Gil, Smichowski, Savio, and Camiña (2017) obtained better results with SIMCA than PLS-DA in classifying different milk-based infant formula according to their elemental composition.

Table **4.2:** Multi-class SIMCA classification[5] of FMP types using the models developed with the full-spectrum and those with the sensitive spectra wavelength range with or without pre-treatment

Model with no pretreatment[1]												
	20% vegetable oil inclusion						30% vegetable oil inclusion					
		Test sets[4] assigned into classes						Test sets assigned into classes				
Class[3]	PC	CO	PO	SBO	SFO	E	PC	CO	PO	SBO	SFO	E
CO	2	**100.0**	100.0	100.0	100.0	100.0	4	**100.0**	100.0	100.0	100.0	100.0
PO	2	100.0	**80.0**	100.0	100.0	89.4	2	100.0	**80.0**	100.0	100.0	89.4
SBO	3	100.0	100.0	**100.0**	60.0	93.1	4	100.0	100.0	**80.0**	80.0	86.4
SFO	2	100.0	100.0	80.0	**40.0**	61.1	3	100.0	100.0	60.0	**80.0**	83.3
Model with pretreatment[2]												
CO	2	**100.0**	100.0	100.0	100.0	100.0	3	**100.0**	100.0	100.0	100.0	100.0
PO	2	100.0	**100.0**	100.0	100.0	100.0	3	100.0	**100.0**	100.0	100.0	100.0
SBO	4	100.0	100.0	**100.0**	100.0	100.0	3	100.0	100.0	**100.0**	40.0	89.4
SFO	4	100.0	100.0	60.0	**60.0**	72.1	4	100.0	100.0	80.0	**60.0**	74.8

[1]No pretreatment performed on the full spectral range, 850 – 2500 nm, of the calibration set (n = 60)

[2]Extended multiplicative signal correction pre-processing was performed only on the sensitive spectra wavelength range, 1216 – 1589 nm, of the calibration set (n = 60)

[3]Model was built for each of the true classes, the FMP types. CO, PO, SBO and SFO represent the coconut oil, palm oil, soya-bean oil and sunflower oil FMP types respectively.

[4]Samples (n = 20) of the FMP types not used for building the models made up the test sets.

[5]PC = optimal number of principal components; E = model efficiency calculated as the geometric mean of sensitivity and specificity. Diagonal elements in bold font represent the sensitivities while the others under the assigned test sets columns indicate the specificity values.

Table **4.3:** Multi-class PLS-DA classification[5] of FMP types using the models developed with the full-spectrum and those with the sensitive spectra wavelength range with or without pre-treatment

Model with no pretreatment[1]												
	20% vegetable oil inclusion						30% vegetable oil inclusion					
		Test sets[4] assigned into classes						Test sets assigned into classes				
Class[3]	LV	CO	PO	SBO	SFO	E	LV	CO	PO	SBO	SFO	E
CO	7	**100.0**	100.0	100.0	100.0	100.0	6	**100.0**	100.0	100.0	100.0	100.0
PO	6	100.0	**100.0**	100.0	100.0	100.0	6	100.0	**100.0**	100.0	100.0	100.0
SBO	10	100.0	100.0	**100.0**	100.0	100.0	13	100.0	100.0	**100.0**	100.0	100.0
SFO	10	100.0	100.0	100.0	**100.0**	100.0	13	100.0	100.0	100.0	**100.0**	100.0
Model with pretreatment[2]												
CO	4	**100.0**	100.0	100.0	100.0	100.0	4	**100.0**	100.0	100.0	100.0	100.0
PO	3	100.0	**100.0**	100.0	100.0	100.0	4	100.0	**100.0**	100.0	100.0	100.0
SBO	9	100.0	100.0	**100.0**	100.0	100.0	12	100.0	100.0	**100.0**	100.0	100.0
SFO	10	100.0	100.0	100.0	**100.0**	100.0	13	100.0	100.0	100.0	**100.0**	100.0

[1]No pre-treatment performed on the full spectral range, 850 – 2500 nm, of the calibration set (n = 60)

[2]Extended multiplicative signal correction pre-processing was performed only on the sensitive spectra wavelength range, 1216 – 1589 nm, of the calibration set (n = 60)

[3]Model was built for each of the true classes, the FMP types. CO, PO, SBO and SFO represent the coconut oil, palm oil, soya-bean oil and sunflower oil FMP types respectively.

[4]Samples (n = 20) of the FMP types not used for building the models made up the test sets.

[5]LV = optimal number of latent variables; E = model efficiency calculated as the geometric mean of sensitivity and specificity. Diagonal elements in bold font represent the sensitivities while the others under the assigned test sets columns indicate the specificities.

4.3.5 Nutritional implications of FMP lipids

The relative ratios and indices necessary for evaluating the nutritional consequences of FMP lipids are shown in Table 4.4. As can be seen, the SFO FMP type recorded the highest values for POA/PO, OA/PA, LA/ALA, and MUFA/SFA. On the other hand, the SBO FMP type recorded the least value (ca. 10.79) for LA/ALA irrespective of the proportions of vegetable oils incorporated. In addition, the PO FMP type had the lowest ratios for POA/PA, OA/PA, and MUFA/SFA. Although there is a lack of literature for FMP to compare these results, it is well documented that commercial infant formulas with a ratio of LA to ALA of 10:1 or higher are nutritionally inadequate (Gibson et al., 1994). The recommended dietary intake of ALA seemed to be about 2 g per day and obtaining an optimal ratio of the two essential FA (i.e., LA and ALA) of less than 4 to 1 in the diet is a major issue (De Lorgeril & Salen, 2004). Thus, to improve the LA/ALA ratio of FMP and justify the fat inclusion on health grounds, two or more vegetable oils should be judiciously selected, keeping in mind their respective LA and ALA contents, and then blended together in optimal proportions. According to Pacheco et al. (2006), the ratio of OA/PA (and that of MUFA/SFA) in dietary fats has a regulatory influence on certain thrombogenic and fibrinolytic markers during postprandial state in healthy subjects. Keys et al. (1986) found in their study that the death rate was negatively correlated to the ratio of MUFA/SFA. Thus, for this study, the SFO FMP type had the best values for OA/PA and MUFA/SFA, but the worst for LA/ALA.

The IA and IT for the FMP types containing SBO and SFO were statistically the same and they recorded the least values. The consumption of diets that are less atherogenic (having low IA) and less thrombogenic (having low IT) may lead to substantial reduction in coronary heart disease (CHD) incidence (Ulbricht & Southgate, 1991). Thus, the SBO and SFO FMP types were the most desirable options in relation to CHD as they had the lowest values, ranging from 0.07 – 0.14 for IA and 0.20 – 0.27 for IT. These values were lower than those (IA = 3.18, IT = 4.11) reported for goat milk by Núñez-Sánchez et al. (2016). They were also lower than those (IA = 1.67, IT = 2.04) reported for whole milk powder by Vargas-Bello-Pérez, Toro-Mujica, Enriquez-Hidalgo, Fellenberg, and Gómez-Cortés (2017). That implies that PO, SBO and SFO FMP types were less atherogenic than whole milk powder. In general, the SBO FMP type had the best overall FA relative ratios and lipid indices for this study and that supported the

recommendation by Modler et al. (1970) to use lightly hydrogenated SBO to formulate FMP with superior health benefits.

Table **4.4:** Relative ratios and lipid indices of the filled milk powder (FMP) types formulated with 3 proportions (10%, 20% and 30%) of 4 different vegetable oils

	Fatty acid relative ratios[2]				Lipid indices[3]	
FMP types[1]	POA/PA	OA/PA	LA/ALA	MUFA/SFA	IA	IT
10% veg. oil inclusion						
CO	-	0.704	-	0.078	14.928[a]	7.227[a]
PO	0.002	0.671	51.590	0.592	1.185[b]	2.416[b]
SBO	0.007	2.045	10.861	1.421	0.138[c]	0.276[c]
SFO	0.016	5.266	281.500	3.073	0.074[c]	0.210[c]
20% veg. oil inclusion						
CO	-	0.642	-	0.023	16.887[a]	8.050[a]
PO	0.002	0.665	52.172	0.589	1.192[b]	2.432[b]
SBO	0.006	2.035	10.758	1.421	0.135[c]	0.271[c]
SFO	0.016	5.006	224.204	2.961	0.072[c]	0.205[c]
30% veg. oil inclusion						
CO	-	0.639	-	0.068	17.059[a]	8.112[a]
PO	0.002	0.660	52.398	0.585	1.199[b]	2.443[b]
SBO	0.006	2.036	10.761	1.425	0.134[c]	0.270[c]
SFO	0.015	5.016	227.405	2.974	0.072[c]	0.204[c]

[a,b,c] Same superscript letters within a column under the same percentage of vegetable oil inclusion are not significantly different at $p \leq 0.05$.
[1]CO, PO, SBO and SFO represent the coconut oil, palm oil, soya-bean oil and sunflower oil FMP types respectively.
[2]POA/PA = palmitoleic to palmitic acid ratio; OA/PA = oleic to palmitic acid ratio; LA/ALA = linoleic to α-linolenic acid ratio; MUFA/SFA = monounsaturated fatty acid to saturated fatty acid ratio.
[3]IA = index of atherogenicity; IT = index of thrombogenicity

4.4 Conclusion

The average near infrared spectral profiles for the FMP types displayed similar absorption characteristics for CO and PO as well as for SBO and SFO making it difficult to differentiate them by visual inspection. The PCA conducted on the full spectra showed the largest variation with respect to the degree of the fat saturation of the FMP samples. It also highlighted the treatment effects by displaying resolved clusters and overlap of samples reflecting the different proportions of vegetable oils incorporated, but without offering substantive separations between most of the individual FMP types. Nevertheless, the SIMCA models were relatively effective in classifying the FMP samples as they recorded mean efficiency values of 92.0 %. However, the PLS-DA models were better classifiers as they achieved mean efficiency values of 100.0 %, thus confirming the existence of significant variances among the FMP types.

In general, the vegetable oils lacked chemical markers that can be used for their immediate identification and classification in FMP samples by FA profile analysis. The SBO FMP type had the best overall FA relative ratios and lipid indices for this study but we recommend that two or more vegetable oils be judiciously selected and blended together in optimal proportions to formulate FMP with superior health benefits. The spectral analytical models we describe may be useful, rapid and cost-effective means by which quality assurance of FMP products can be monitored.

Chapter 5

Paper III

Effective detection and quantification of chemical adulterants in model fat-filled milk powders using NIRS and hierarchical modelling strategies

Abstract

Skimmed milk powder can be blended with indigenous vegetable oils to formulate fat-filled milk powder (FMP) to satisfy the growing demand of low-income consumers for dairy proteins in the developing countries where food control systems may be fragmented. Unscrupulous manufacturers may adulterate FMP with melamine or urea to give the false impression that it contains sufficient protein. This study investigated, for the first time, the efficacy of near infrared spectroscopy to detect and quantify melamine and urea (0.01 – 16.00%) in FMP formulated with 4 different vegetable oils (i.e., coconut, palm, soya-bean and sunflower). Multilevel analyses were able to detect, confirm and differentiate the adulterations with an efficiency ranging from 89.8 to 100.0%. The partial least square regression models yielded satisfactory predictions ($R^2_p \geq 0.96$, RSR ≤ 0.19) at adulteration levels ≥ 1.00%. This study provides appropriate guidelines for cost-efficient screening of FMP products for adulterants to protect public health.

Chemical compounds studied in this chapter:
Urea (PubChem CID: 1176)
1,3,5-triazine-2,4,6-triamine (PubChem CID: 7955)

keywords: fat-filled milk powder, melamine and urea adulteration, near infrared spectroscopy, hierarchical modelling strategy, multilevel SIMCA screening, interval PLS

5.1 Introduction

Fat-filled powders (FMP) are produced by blending skimmed milk powder (SMP) with vegetable fat and the standards for their formulation are contained in the regulations of the Codex Alimentarius Commission of the Food and Agriculture Organisation and World Health Organisation adopted in 2006 (last amended in 2014) (Codex Alimentarius Commission, 2006). It specifies the minimum amount of milk protein in milk solids-not-fat (i.e., 34 % w/w) that must be met when replacing milk fat with a locally available vegetable oil to produce an affordable alternative to whole milk powder (Codex Alimentarius Commission, 2006). FMP is exported in bulk and then packed into small/single serve portions (European Commission, 2017) to satisfy the growing demand of the low-income consumers for dairy proteins in developing markets, including many African nations where the border controls are poor and the national food control systems are fragmented (FAO/WHO, 2003). Consequently, the proof of safety is left to the manufacturers; unfortunately, unscrupulous producers or their intermediaries may capitalise on the opportunity to tamper with the natural constituents of FMP (i.e. the milk proteins) for the purpose of increasing its apparent value for financial gains (Handford et al., 2016). The rising sales for FMP (e.g., African imports grew by 8 % per year between 2007 and 2017 (European Commission, 2017)) provide a driver for fraudulent activity (Handford et al., 2016). Hence, the low-cost melamine (1,3,5-triazine-2,4,6-triamine) or urea, with no nutritional value, may be deliberately added to FMP to give the false impression that it contained high protein levels. It has long been known (Pei et al., 2011; Xiu & Klein, 2010) that the ingestion of melamine (a noxious nitrogen-rich chemical normally used in plastics) may cause serious damage to internal organs especially in young children. Thus, while the primary motivation for the fraudulent addition of prohibited substances into FMP is to maximize profit for the few, the impact is a real threat to public health and safety of the many. The extent of such adulteration to developing nations is substantive (Handford et al., 2016).

The potential for tampering is a strong driver for a rapid, inexpensive method for screening and quantifying adulterants in FMP. Several researchers (Botros et al., 2013; Capuano et al., 2015; Chen, Tan, Lin, & Wu, 2017) have successfully used spectroscopic methods such as near infrared spectroscopy (NIRS) for profiling the quality and safety of milk products because they often require minimal or no sample preparation, as well as providing rapid (one minute or less per sample) and precise on-line analysis, with the potential for running multiple tests on a

single sample (Nawrocka & Lamorska, 2013). To the best of our knowledge, NIRS has never been used for screening adulterants in FMP, although its applicability might be challenged by the complexity of the sample matrices resulting from the mixture of the backbone fats (i.e., the vegetable lipids) and the adulterants (if any) in the fat-filled milk powder. However, it has been reported (Craig, Botelho, Oliveira, & Franca, 2018; De Souza Gondim, Junqueira, De Souza, Ruisánchez, & Callao, 2017; Reis, Botelho, Franca, & Oliveira, 2017) that the use of hierarchical modelling strategy (HMS) can minimise the error rates and improve the model performance when the sample variability of a spectroscopic data is high. Furthermore, we have previously described the validity of a NIRS-HMS strategy for the differentiation of model FMP formulated with different fat species (Ejeahalaka & On, 2019a). In this study, chemometric techniques were applied using HMS as a tool to qualify and quantify melamine and urea adulteration in the near infrared spectral profiles of FMP produced onsite with different backbone fats and spiked with differing concentrations of the adulterants.

5.2 Material and methods

5.2.1 Raw materials collection

The four vegetable oils selected for this study contained high (e.g., coconut oil), medium (e.g., palm oil) and low (e.g., soya-bean oil and sunflower oil) concentrations of saturated fatty acids (SFA). Palm oil (PO) was purchased from Pure Nature (New Zealand) whereas coconut oil (CO), soya-bean oil (SBO), sunflower oil (SFO) and SMP were purchased from a local supermarket. Melamine and urea of appropriate analytical grade (purity ≥ 99 %) were purchased from Sigma-Aldrich, Christchurch in New Zealand. Both melamine and urea have been used to adulterate milk powders (Handford et al., 2016).

5.2.2 Model fat-filled milk powder manufacture and adulteration

Eighty pure (unadulterated) FMP samples were formulated onsite at the analytical laboratory of Lincoln University using the methods described previously (Ejeahalaka & On, 2019a). In brief, 25 g of SMP was first dispersed and mixed in 250 mL of deionised water at 20 °C using a magnetic stir bar while avoiding the incorporation of air. Then, the four vegetable oils (i.e., CO, PO, SBO and SFO) used for the formulation were pre-warmed to 15 ^{o}C above their respective melting points before 30 % of each one of them (minimum permitted total fat in FMP is 26 % w/w (Codex Alimentarius Commission, 2006)) was independently added to the

reconstituted SMP. The mixtures were then homogenised and lyophilised resulting in a total of 20 samples for each of the 4 FMP types.

For the adulterated samples, the blank FMP was first exposed to laboratory conditions (temperature (25° C) and relative humidity (50 %)) for 2 hours in airtight containers and allowed to equilibrate. Then, it was partially substituted with either melamine or urea independently (w/w) to 10 concentration levels (0.01 %, 0.04 %, 0.08 %, 0.20 %, 0.60 %, 1.00 %, 4.00 %, 8.00 %, 12.00 % and 16.00 %) ranging from low to high in sealed polypropylene tubes in triplicates to achieve a final powder blend of 10 g in each case. The dried mixture blends were manually homogenised by repeated inversion for 20 times to ensure thorough mixing and then stored at room temperature until the NIRS analysis. A total of 30 adulterated fat-filled milk powder (AFMP) samples was obtained for all the concentration levels per milk type resulting in 120 samples for each adulterant (i.e., melamine or urea). CO, PO, SBO and SFO FMP types adulterated with melamine were denoted as COM, POM, SBOM and SFOM whereas those adulterated with urea were COU, POU, SBOU, and SFOU respectively. Measurements were carried out within a very short span after sample preparation in order to minimise experimental errors (Balabin & Smirnov, 2011).

5.2.3 Near infrared spectral acquisition

The near infrared spectra of the model (unadulterated) FMP and that of the FMP types spiked with either melamine or urea were obtained using a FOSS NIRSystem (model DS 2500F) spectrometer as described previously (Ejeahalaka & On, 2019a). Briefly, 5 g of each sample placed in a ring cup (type 2004) was scanned at a resolution of 0.5 nm, wavelength accuracy of < 0.05 nm, analysis time of < 1 minute and a detector array composed of both silicon (850 – 1100 nm) and lead sulfide (1100 – 2500 nm). A total of 319 unique samples (i.e., 20 from each of the 4 unadulterated FMP types, 119 from FMP types adulterated with melamine (as one sample was contaminated and as such discarded) and 120 for those adulterated with urea) were scanned in triplicates in two consecutive days of sampling without shutting down the machine. Each spectrum comprised of 3300 absorbance values recorded from 850 to 2499.5 nm averaging 32 scans. The triplicate spectra acquired from each of the samples were then averaged and subjected to chemometric modelling.

5.2.4 Data analysis and chemometric modelling

The raw spectra of the 319 unique samples were transferred into the R software version 3.5.1 (R Core Team, 2018) and analysed using the principal component analysis (PCA) technique to identify any natural groupings within the spectral data. Raw spectra were pre-processed using the second derivative Savitzky-Golay algorithm (Savitzky & Golay, 1964)(window size = 25 points, second-order polynomial fit) followed by the extended multiplicative signal correction (Stark & Martens, 1996) to reduce the noise level and the scattering effects to enhance the signal properties. Thereafter, pre-processed spectra were randomly divided into a training set (to build the model) and a testing set (to test the model) in the ratio of 4:1.

5.2.4.1 Classification by hierarchical modelling strategy (HMS)

Implementation of HMS was achieved by building a sequence of simpler models based on targeted and untargeted approaches instead of one multiclass model containing all the studied classes (Craig et al., 2018). Consequently, three levels of soft independent modelling of class analogy (SIMCA) models were built based on HMS to predict the class memberships of all the FMP samples (adulterated and unadulterated).

In the first level, a one-class SIMCA model was built using the training set of all unadulterated FMP samples (i.e., CO, PO, SBO and SFO) combined together irrespective of milk type. The model was based on an untargeted approach to predict whether the unknown test sample presented to it for screening was unadulterated (where a fit was found) or adulterated (where a fit was not found). In the second level, four one-class SIMCA models were built for all unadulterated FMP types (as above) based on a targeted approach to test the ability of each one to recognise the test samples that belong to their individual classes and those that are adulterated with melamine or urea. Class models were further validated using each pure adulterant and limits to which melamine and urea can be detected in each of the four FMP types also quantified at this level.

In the third level of the hierarchy, adulterant-specific models were built based on a targeted approach to predict the milk classes of the adulterated test samples which were predicted to contain either melamine or urea in level two. Eight one-class SIMCA models were built in the third level with the FMP types adulterated with melamine (i.e., COM, POM, SBOM, and SFOM) or urea (i.e., COU, POU, SBOU, and SFOU) for the identification and differentiation of the test

samples. Hence, different possible combinations of the AFMP test sets were presented to the class models to test their abilities to detect them.

To implement the SIMCA models for the three levels defined above, the optimal number of principal components (PC) was determined through internal validation using the leave-one out (LOO) cross-validation method (S. Wold, 1978). The calibration models were built using the full spectra range (850 – 2500nm, 3300 wavelengths) and then optimised using only the selected sensitive wavelength range. To do this, the variables that carried most of the information were first identified from the PCA loadings and then, an optimal interval was found around them to build the models. Model performance was assessed based on sensitivity (i.e., the ability to screen in correctly samples belonging to the modelled class), specificity (i.e., the ability to screen out correctly samples not belonging to the modelled class), efficiency, and reliability rate. The efficiency (E) of the model was calculated as the geometric mean of the values of sensitivity and specificity (Chong & Jun, 2005; Oliveri & Downey, 2012). The reliability rate (RR) was calculated according to Reis et al. (2017) to provide a better overview of the trueness of the models. Details of algorithms used are given in Appendix A (Section A.6).

5.2.4.2 Quantitative prediction of adulterants in FMP

Partial least square regression (PLSR) was employed as the supervised multivariate technique to correlate the spectral intensities of the AFMP samples to the chemical adulterants (i.e., melamine and urea) concentrations across the 3300 wavelengths. A PLSR model was built for each of the 8 AFMP types using 23 samples as the training set and predictions were made on 7 independent test set to evaluate the predictive ability of the model. The optimal number of latent variables (LV) was determined by LOO cross-validation. To find out the extent/limit to which the chemometric models can correctly predict the adulterants in the test samples for each of the AFMP types, the relative error at each of the predicted concentration levels was calculated as $(Mi - Pi)/Mi$ where M_i and P_i were the measured and the predicted values for the i^{th} sample respectively. Each prediction was considered satisfactory/acceptable if the relative error between the measured concentration of the adulterant and the predicted value was less than 0.20. The overall capability of the models was determined using the root mean squared error of prediction (RMSEP) (Porep, Kammerer, & Carle, 2015), RMSE-observations standard deviation ratio (RSR) (Moriasi et al., 2007) and the percent bias (PBIAS) (Moriasi et al., 2007). In addition, PRESS R^2 denoted in this study as R^2_p, which is a measure of the

predictive performance of the models (Oftedal, Eisert, & Barrell, 2014), was used to assess the agreement of the prediction results with the measured concentrations.

PLSR predictions obtained in the present study using the full spectrum (850 – 2500nm, 3300 wavelengths) were optimised using the interval PLS (iPLS) variable selection technique (Nørgaard et al., 2000) to find one or a few intervals that would give better predictions than the predictions obtained when using the full spectrum, and to provide an overview that would aid in spectroscopic interpretations (Andersen & Bro, 2010). Details of algorithms used are given in Appendix A (Section A.6).

5.3 Results and discussion

5.3.1 Characteristics of unadulterated and adulterated FMP NIRS profiles

Figure 5.1 shows the representative raw near infrared spectra of the unadulterated FMP types (CO, PO, SBO, and SFO) and those adulterated with 16 % of either melamine (COM, POM, SBOM and SFOM) or urea (COU, POU, SBOU, and SFOU). The profile analysis for all the proportions of adulterations from 0.01 % up to 16.00 %, in addition to those for the pure adulterants, are shown in Appendix A (Fig. A2 to Fig. A5). As can be seen from the figures, the profiles of the pure (100 %) adulterants (melamine or urea) were clearly distinguishable from those of the FMP types. However, all milk spectra (adulterated or unadulterated) had the characteristic FMP strong absorption peaks at about 1210 nm, 1726 nm and 2350 nm which increased in intensity and broadness with increasing unsaturation of the incorporated vegetable oil (Ejeahalaka & On, 2019a). The spectral profiles for the adulterated FMP looked similar in every respect to that of the unadulterated FMP types at adulteration level ≤ 1 % (Fig. A2 to Fig. A5) except for the visible differences in their absorption intensities. That implies that although the dry-blending of 1 % (w/w) of melamine (or urea) with FMP increased the nitrogen and thus the apparent protein content by about 46 % (or 43 % in the case of urea) (Table A1) as measured by Dumas method (Oftedal et al., 2014), the spectral differences (between adulterated and unadulterated) generally appeared indiscernible to the naked eye. However, a close inspection revealed that there was a significant separation between the spectral profiles of the unadulterated and the adulterated CO FMP types. The addition of melamine or urea to CO FMP caused a noticeable increase in absorbance values across the entire wavelengths, thus highlighting the high specificity of coconut oil to either of the chemical adulterants. On the other hand, there was a spectral overlap between the unadulterated and

the adulterated PO FMP types, and that probably highlighted the ability of palm oil to mask the presence of the chemical adulterants. Overall, the structural effects of the progressive addition of melamine and urea to FMP became apparent at concentration levels ≥ 4 % (Fig. A2(b)(d), Fig. A3(b)(d), Fig. A4(b)(d) and Fig. A5(b)(d)) mainly around the regions near 1465 to 1530 nm and 1955 to 2150 nm. Consequently, all the spiked samples (at concentration level ≥ 4 %) exhibited the prominent melamine and urea strong absorption peak at about 2000 nm band arising from the N-H stretching/N-H deformation combination (Abbas, Lecler, Dardenne, & Baeten, 2013). The melamine adulterated samples had a unique band at about 1955 nm corresponding to the first overtone of NH stretching (Abbas et al., 2013) as compared to those adulterated with urea which also had a relatively definable unique absorption band near 1315 nm. However, the samples containing urea generally had lower absorbance values and the adulterated PO FMP types recorded the least intensities. The order of the absorption intensities of the adulterated FMP types was as follows:

SBO > SFO > CO > PO.

This ordering pattern is significantly different from that of the unadulterated FMP types that was previously reported (Ejeahalaka & On, 2019a), thus highlighting the possibility of some compositional alterations in the sample matrices as a result of the addition of the chemical adulterants.

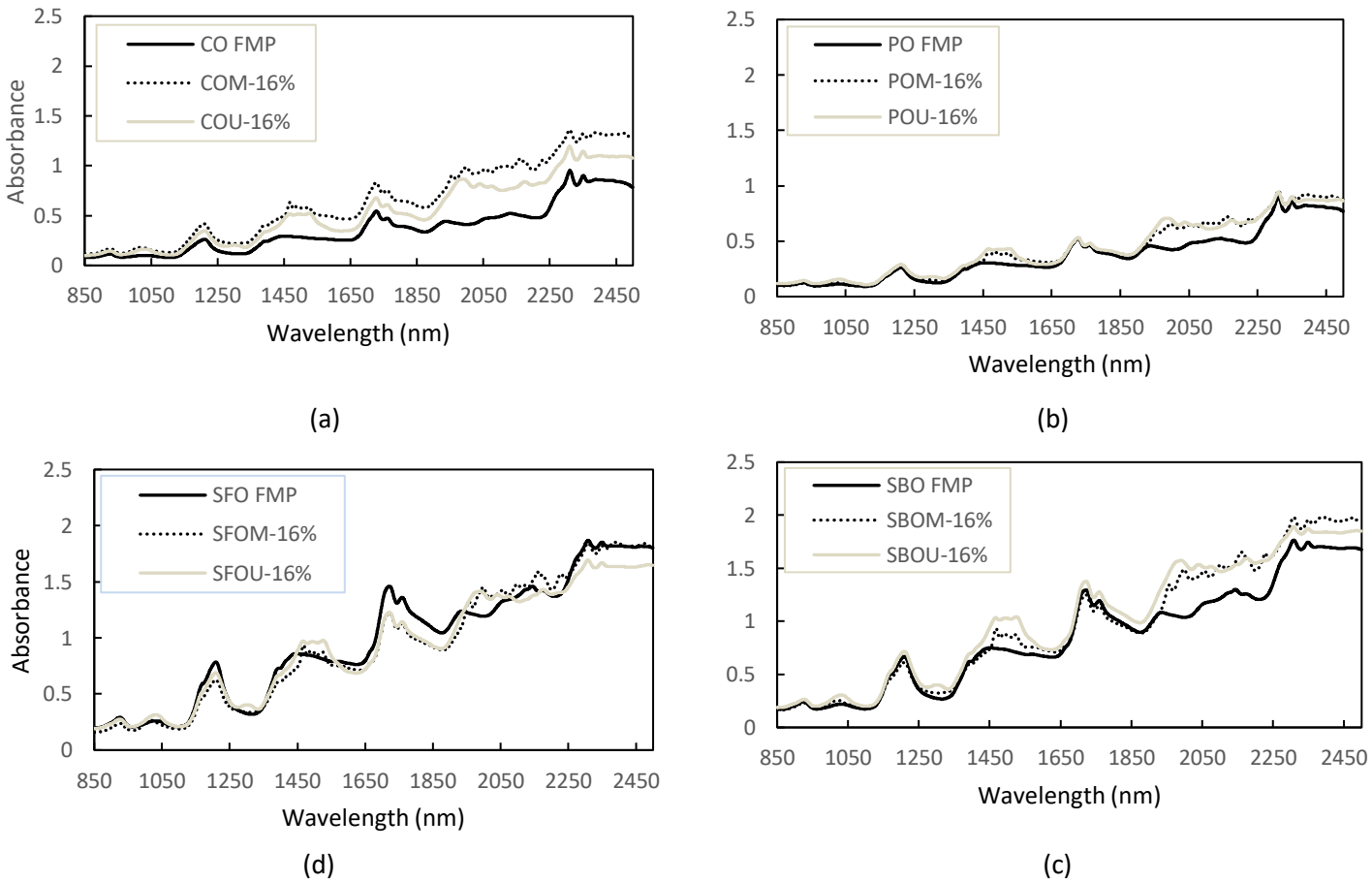


Figure **5.1:** Representative spectra profiles of unadulterated and adulterated (a) coconut oil FMP (b) palm oil FMP (c) soya-bean oil FMP and d) sunflower oil FMP. CO, COM, COU, PO, POM, POU, SBO, SBOM, SBOU, SFO, SFOM, and SFOU represent FMP types containing coconut oil, coconut oil with melamine, coconut oil with urea, palm oil, palm oil with melamine, palm oil with urea, soya-bean oil, soya-bean oil with melamine, soya-bean oil with urea, sunflower oil, sunflower oil with melamine and sunflower oil with urea respectively.

5.3.2 PCA adulterated and unadulterated fat-filled milk powder characteristics

The pattern of similarity of the FMP samples (adulterated and unadulterated) and of the 3300 near infrared spectra variables was explored by PCA. The results obtained from the PCA analysis were summarised in the score plot depicted in Figure 5.2. The dimensionality of the spectral data was reduced by projecting the samples onto the first two principal components (PC 1) and (PC 2) which retained 99 % of the original variance in the 3300 spectral variables. It can be seen from the score plot that PC 1 was effective in separating the adulterated and the unadulterated CO and PO FMP samples from those of SBO and SFO FMP types in line with their degree of saturation in lipids. All the adulterated and unadulterated PO FMP samples clustered closely together with those of the unadulterated CO FMP type but were separated from those of CO adulterated with melamine or urea. This result supported the mean spectral profile analysis in highlighting the contrasting abilities of palm oil and coconut oil in masking the presence of the studied chemical adulterants. However, the PCA did not show a cohesive clustering between the adulterated and the unadulterated SBO and SFO FMP types, indicating that this analysis was unsuited for providing insight into their sample histories.

The loadings variable plot (Figure 5.3) showed that 11 wavelengths (i.e., 1020, 1216, 1310, 1425, 1465, 1530, 1713, 2000, 2130, 2155 and 2225 nm) contributed to the sample separation in the mean centred PCA of the adulterated and the unadulterated FMP types. The peaks and troughs of the loadings for PC 1 and PC 2 were at the same wavelengths at 1216 and 1713 nm which most likely correlated with the C-H bond absorption (Workman & Weyer, 2012) of vegetable fats. The wavelengths at 1020, 1465 and 1530, 2000, 2155 and 2225 nm may be related to N-H stretching second overtone, N-H stretching first overtone, N-H stretching/N-H deformation combination, and 1,3,5-triazine structural vibrations respectively (Abbas et al., 2013). Thus, the two adulterants (melamine and urea) contributed largely to PC 2 loadings plot, with the wavelength intervals, 1465 - 1530 nm and 2000 – 2225 nm, differentiating the spectral profiles of the adulterated and the unadulterated FMP types. However, the band, 1465 to 1530 nm, which was previously described as the characteristic melamine spectral features (Scholl et al., 2017), was used as the most discriminative wavelength range for building the optimised models as they yielded more stable and better predictions.

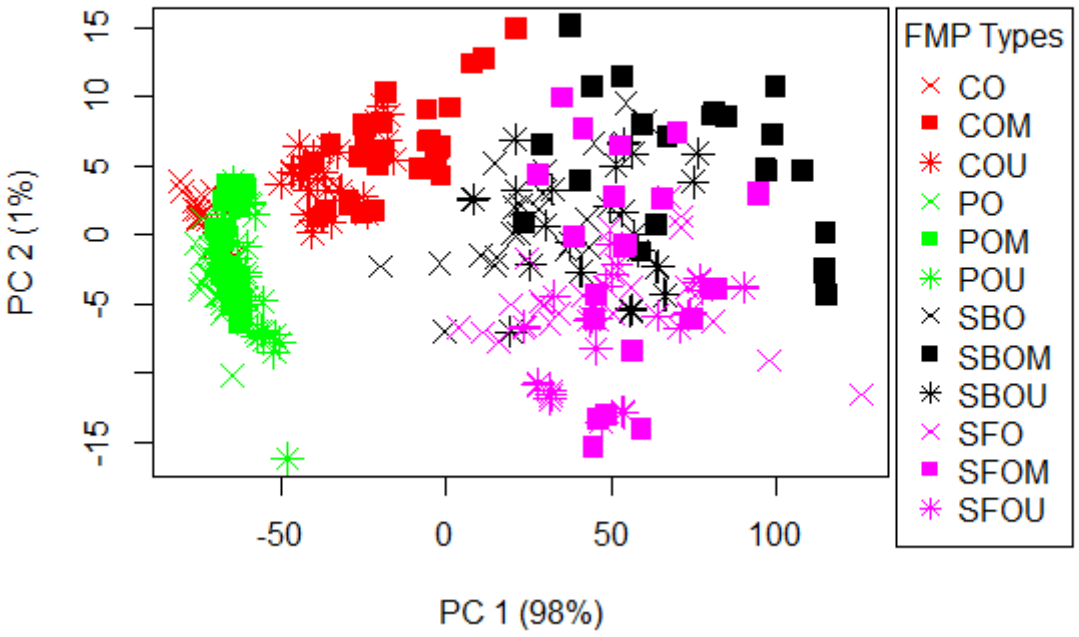


Figure **5.2:** PCA score plot of typical and atypical FMP types. CO, COM, COU, PO, POM, POU, SBO, SBOM, SBOU, SFO, SFON, and SFOU represent FMP types containing coconut oil, coconut oil with melamine, coconut oil with urea, palm oil, palm oil with melamine, palm oil with urea,

soya-bean oil, soya-bean oil with melamine, soya-bean oil with urea, sunflower oil, sunflower oil with melamine and sunflower oil with urea respectively.

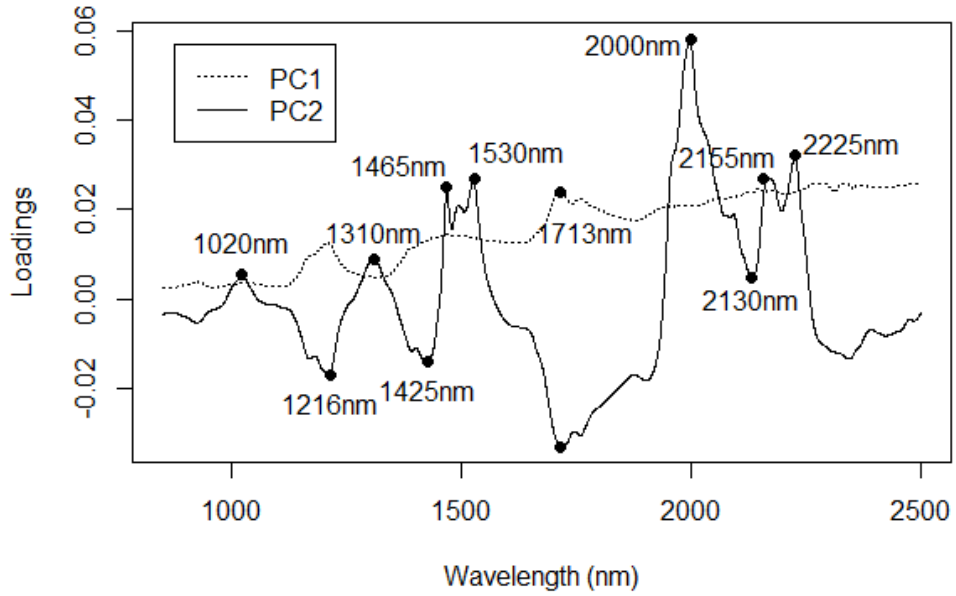


Figure **5.3:** Loadings variable plot of the first two principal components, PC 1 and PC 2, of mean-centred PCA of the typical and atypical FMP types showing the sensitive wavelengths.

5.3.3 Multilevel chemometric detection of chemical adulterants in FMP

Three multilevel SIMCA models were developed to reduce complexity, improve interpretability and to account for the shared variance at different levels of the spectral hierarchy. The training sets were used for building the classification rules and for quantifying the adulterants in the FMP samples whereas the testing sets was used as an independent validation data to determine the accuracy of the predictions. The first HMS level model was built using the calibration sets (n = 60) of all the unadulterated samples combined together to determine its ability to detect whether the unknown test sample presented to it for screening was typical or atypical. It can be seen from Table 5.1 that the one-class SIMCA model classified the objects with low false negatives (3 of 20) and low false positives (22 of 239), yielding sensitivity and specificity values of 85.0 % and 90.8 % respectively. This shows that it is possible to build a SIMCA model with an acceptable recognition and rejection rates for the FMP samples without any reference to the types/proportions of either the incorporated vegetable

oils or the spiked chemical adulterants. Hence, the first level model was able to detect the adulterations in the FMP samples with high specificity.

The second HMS level model was built as a follow up to the first level using the calibration set (n = 15) for each of the four unadulterated FMP types to test their abilities to detect whether the unknown test sample belonging to the modelled class was unadulterated or adulterated with either melamine or urea. Consequently, four models were built, one for each unadulterated FMP type, and they were tested using all the melamine (n = 119) or urea (n = 120) adulterated samples combined together irrespective of milk type and adulteration levels. Two types of treatments were presented namely: full-spectrum (i.e., 850 – 2500 nm, 3300 wavelengths) models and optimised models (built with selected sensitive wavelength range, 1465 – 1530 nm, 131 wavelengths). As can be seen from Table 5.1, the optimised models performed better than the full-spectrum models and as such will only be discussed forthwith. The one class SIMCA model for CO FMP was clearly able to recognise its members that were unadulterated and those of all the milk types adulterated with melamine or urea as it recorded maximum (100.0 %) sensitivity and specificity values. That implies that the limit of detection (LOD) of the CO FMP model for melamine or urea was at least 0.01 % since that was the minimum concentration level of the chemical adulterants for this present study. This result supported the PCA and the mean spectral profile analyses in highlighting the high specificity of coconut oil FMP to melamine and urea adulteration. For the PO FMP type, the sensitivity was 100.0 % and the specificity values for melamine and urea adulterations were 98.3 % and 100.0 % respectively. The melamine adulteration had 2 (out of 119) false positives and the concentration level recorded (data not shown) for these wrongly accepted adulterated samples was 0.01 %. Thus, the LOD of the PO FMP model for melamine or urea was as low as 0.01 %, highlighting the possibility of uncovering the masking ability of palm oil for the studied chemical adulterants using the power of chemometric analysis. The one-class SIMCA model for the SBO FMP type had a sensitivity of 100.0 % and specificity values of 100.0 % and 93.3 % for melamine and urea adulterations respectively. The concentration levels for the 8 false positives recorded for the urea adulterations were ≤ 0.04 %. Thus, the LOD of the SBO FMP model for melamine was lower than that for urea adulteration. On the other hand, the specificity of the one-class SIMCA model for SFO FMP type for melamine and urea were 100.0 % and 90.0 % respectively. The false positives were 12 of 120 for urea adulteration and those samples had concentration levels ≤ 0.08 %. Thus, the LOD of the SFO FMP model for urea

adulteration exceeded 0.08 % whereas that for melamine was at least 0.01 %. Thus, urea was detected with higher false positives in FMP types formulated with vegetable oils of higher unsaturation in lipids. Overall, all the typical FMP types had LOD of at least 0.01 % for melamine and LOD ranging from 0.01 % to 0.20 % for urea adulterations. Hence, the second level model was able to confirm the types of adulteration of the FMP samples and their limits of detection. These results suggest that although the presence of the studied chemical adulterants can be effectively detected in FMP using NIRS, the LOD that can be achieved may still be higher than the threshold of 0.0001 % (1 ppm) set by the Codex Alimentarius Commission (2013) for infant milk powder. Although we have found no previous studies to detect melamine or urea in FMP, melamine has previously been detected in milk powder (Balabin & Smirnov, 2011) at LOD < 1 ppm using NIRS, supporting the use of this technology for low-level melamine detection in FMP.

The third HMS level model was built to identify and differentiate the atypical FMP samples which were found to be adulterated with either melamine or urea in the second level modelling. Hence, 8 adulterant-specific models were built, one for each FMP type adulterated with either melamine or urea, using 23 samples as the calibration set. The model performance metrics are shown in Table 5.2. The optimised models (1465 – 1530 nm, 131 wavelengths) performed better than the full-spectrum models (850 – 2500 nm, 3300 wavelengths) as they recorded higher mean reliability rate and efficiency of 94.2 % and 97.1 % respectively. Thus, only the results for the optimised models will be discussed here. As can be seen from the table, the main diagonal elements represent the sensitivity values whereas the off-diagonal elements summarise their specificities. All the models had sensitivities ≥ 85.7 % and such values had been shown (De Souza Gondim et al., 2017) to be high enough for them to recognise their own members. On the other hand, 7 out of the 8 models had specificity values ≥ 71.4 % indicating that they can successfully differentiate between the adulterated FMP types. SBOU was the only model which recorded low specificity value of 57.1 % while differentiating the adulterated test samples from the SFOU FMP type. This model presented high false positives because of the similarities in sample matrices between SBO and SFO FMP types making them difficult to differentiate. The minimum and maximum reliability rates recorded across the adulterated FMP models were 79.9 % and 100.0 % respectively. The one-class SIMCA models for COM and POM were the most efficient (100.0 %) in identifying and differentiating the adulterated objects whereas the SBOU FMP model was the least (89.8 %).

Table **5.1:** Sensitivity and specificity values for the first[1] and second[2] levels multi-class SIMCA models for unadulterated and adulterated FMP types[3] using the Hierarchical Modelling strategies

First Level Model			
	Predicted unadulterated	Predicted adulterated	
True class (unadulterated)	17/20 (85.0 %)	217/239 (90.8 %)	
Second Level Model			
		Predicted adulterated	
True class (unadulterated)	Predicted unadulterated	Melamine	Urea
SG + EMSC Full[4]			
CO	5/5 (100.0 %)	119/119 (100.0 %)	120/120 (100.0 %)
PO	4/5 (80.0 %)	111/119 (93.3 %)	120/120 (100.0 %)
SBO	4/5 (80.0 %)	119/119 (100.0 %)	115/120 (95.8 %)
SFO	4/5 (80.0 %)	108/119 (90.8 %)	103/120 (85.8 %)
SG + EMSC Var[5]			
CO	5/5 (100.0 %)	119/119 (100.0 %)	120/120 (100.0 %)
PO	5/5 (100.0 %)	117/119 (98.3 %)	120/120 (100.0 %)
SBO	5/5 (100.0 %)	119/119 (100.0 %)	112/120 (93.3 %)
SFO	4/5 (80.0 %)	119/119 (100.0 %)	108/120 (90.0 %)

[1]First level model checks whether the unknown samples (n = 259) are typical or atypical.

[2]Second level model predicts the type of the adulterated and the unadulterated test samples.

[3]CO, PO, SBO and SFO represents coconut oil, palm oil, soya-bean oil and sunflower oil fat-filled milk powder (FMP) types respectively.

[4]SG + EMSC Full = Savitzky-Golay (second derivative with second polynomial order and a window size of 25 nm) plus the extended multiplicative signal correction pre-processing methods performed on the full spectral range, 850 – 2500 nm of the calibration sets (n = 60).

[5]SG + EMSC Var = Savitzky-Golay (second derivative with second polynomial order and a window size of 25 nm) plus the extended multiplicative signal correction pre-processing methods performed only on sensitive bands from 1465 to 1530 nm of the calibration sets (n = 60).

Table **5.2:** Performance metrics for the third level multi-class SIMCA models for the adulterated FMP types[1] using the Hierarchical Modelling strategies

Third Level Model – Adulterants Identification and Differentiation										
	Sensitivity[2] and specificity (%)								Robustness[3]	
True Class	COM	COU	POM	POU	SBOM	SBOU	SFOM	SFOU	RR	E
SG + EMSC Full[4]										
COM	**85.7**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	85.7	92.6
COU	100.0	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
POM	100.0	100.0	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0
POU	100.0	100.0	42.9	**100.0**	100.0	100.0	100.0	100.0	92.2	96.0
SBOM	100.0	100.0	100.0	100.0	**71.4**	57.1	87.5	100.0	63.7	81.2
SBOU	100.0	100.0	100.0	100.0	100.0	**71.4**	100.0	100.0	71.4	84.5
SFOM	100.0	100.0	100.0	100.0	100.0	100.0	**100.0**	100.0	100.0	100.0
SFOU	100.0	100.0	100.0	100.0	100.0	100.0	100.0	**71.4**	71.4	84.5
SG + EMSC Var[5]										
COM	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
COU	100.0	**87.5**	100.0	100.0	100.0	100.0	100.0	100.0	87.5	93.5
POM	100.0	100.0	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0	100.0
POU	100.0	100.0	71.4	**100.0**	100.0	100.0	100.0	100.0	96.1	98.0
SBOM	100.0	100.0	100.0	100.0	**100.0**	85.7	87.5	100.0	96.2	98.1
SBOU	100.0	100.0	100.0	100.0	100.0	**85.7**	100.0	57.1	79.9	89.8
SFOM	100.0	100.0	100.0	100.0	100.0	100.0	**100.0**	85.7	98.0	99.0
SFOU	100.0	100.0	100.0	100.0	100.0	71.4	100.0	**100.0**	96.2	98.1

[1]COM, COU, POM, POU, SBOM, SBOU, SFOM and SFOU represent the fat-filled milk powder (FMP) types containing coconut oil with melamine, coconut oil with urea, palm oil with melamine, palm oil with urea, soya-bean oil with melamine, soya-bean oil with urea, sunflower oil with melamine and sunflower oil with urea respectively.

[2]Sensitivity values for test sets (n = 59) are highlighted across the main diagonal of the table.

[3]RR = reliability rate, E = efficiency and both values are expressed in percent.

[4]SG + EMSC Full = Savitzky-Golay (second derivative with second polynomial order and a window size of 25 nm) plus the extended multiplicative signal correction pre-processing methods performed on the full spectral range, 850 – 2500 nm of the calibration sets (n = 180).

[5]SG + EMSC Var = Savitzky-Golay (second derivative with second polynomial order and a window size of 25 nm) plus the extended multiplicative signal correction pre-processing methods performed only on sensitive bands from 1465 to 1530 nm of the calibration sets (n = 180).

5.3.4 PLSR modelling for the quantification of chemical adulterants in FMP

Good quantitative calibration models are expected to have low PBIAS, RSR < 0.5 and RMSEP of value less than half the standard deviation of the measured data (Moriasi et al., 2007). In addition, acceptable predictive models are expected to have low relative error and $R^2_p > 0.95$ (Oftedal et al., 2014).

The performance statistics for the PLSR models developed for quantifying the melamine and urea adulterations in FMP types are shown in Table 5.3. As can be seen from the table, the root mean squared error of cross-validation (RMSECV) for the iPLS optimised models were 27 – 92 % lower than those obtained for the full-spectrum models. Also, the optimised models had a greater number of predictions with acceptable relative errors (< 0.20) and a maximum of 5 latent variables was retained for the predictions in order to achieve an optimal generalization (Abdi, 2010). Thus, only the iPLS optimised models will be discussed here. All the models showed excellent results ($R^2_p > 0.96$, RSR ≤ 0.19) with the SFOU FMP type having the best RSR value of 0.03 (optimal < 0.50). A small maximum difference of 0.44 % (0.97 % for the full-spectrum models) was found between RMSECV and RMSEP and that indicated that the models were robust enough for both the calibrations and the predictions. However, the models for COU, POU and SBOU FMP types had the largest differences between the two error values, RMSECV and RMSEP. The COU FMP type recorded the maximum RMSEP of 1.12 % which was less than the 2.99 % value calculated as half the standard deviation of the measured concentration levels that was desired for good quantitative calibration models. A close inspection of the relative errors at 6 concentration levels revealed that most of the predictions were satisfactory at 1.00 % adulteration level where the relative errors between the measured concentrations of the adulterants and the predicted values were less than 0.20. However, the relative errors increased dramatically at concentration levels < 1.00 % resulting in a loss of accuracy of quantification of the prediction models. In general, the melamine adulterated FMP types were better predicted than those of urea as they mostly had lower RSR values.

Table **5.3:** Performance statistics[1] of PLS regression models for estimating the concentration levels of melamine and urea adulterations[2] in fat-filled milk powder using the full spectral wavelengths and the interval PLS variable selection

			Prediction								
	Calibration		Relative error[5] at each spiked concentration level								
AFMP	LV	RMSECV	0.08%	0.20%	1.00%	4.00%	8.00%	16.00%	R^2_p	RMSEP	RSR
SG + EMSC Full[3]											
COM	5	0.67	0.38	2.22	**-0.17**	**-0.05**	**0.00**	**0.05**	1.00	0.36	0.06
COU	4	2.09	-18.4	7.52	**-0.08**	**0.06**	**-0.14**	**-0.02**	0.96	1.12	0.19
POM	2	0.36	3.03	-1.21	**0.08**	**-0.04**	**-0.04**	**-0.01**	1.00	0.19	0.03
POU	2	0.31	0.38	-0.41	**0.01**	**0.01**	**-0.02**	**-0.12**	0.98	0.75	0.13
SBOM	2	1.30	8.71	1.66	-0.23	-0.32	**-0.01**	**0.10**	0.98	0.94	0.14
SBOU	7	0.69	2.17	-3.71	**-0.03**	**0.02**	**0.15**	**-0.01**	0.99	0.55	0.09
SFOM	2	1.18	8.43	2.34	-0.33	-0.21	**-0.09**	**0.15**	0.96	1.09	0.18
SFOU	6	0.49	4.10	3.14	0.43	**0.19**	**-0.05**	**0.04**	0.99	0.55	0.09
SG + EMSC iPLS[4]											
COM	5	0.40	-1.52	**-0.18**	**-0.16**	**0.05**	**0.05**	**0.06**	0.99	0.43	0.07
COU	5	1.56	-4.94	**0.04**	0.35	**-0.01**	**-0.12**	**0.04**	0.96	1.12	0.19
POM	2	0.33	2.37	-0.80	**-0.04**	**-0.06**	**-0.05**	**-0.01**	1.00	0.22	0.04
POU	2	0.27	**-0.16**	-0.47	**0.08**	**0.04**	**-0.01**	**-0.11**	0.99	0.64	0.11
SBOM	5	0.35	-7.24	**-0.14**	**-0.03**	0.24	**-0.05**	**0.00**	0.99	0.48	0.07
SBOU	5	0.29	-1.31	-1.92	**-0.14**	**0.11**	**0.15**	**0.01**	0.99	0.54	0.09
SFOM	3	0.44	**0.07**	0.55	**-0.06**	**-0.09**	**-0.12**	**0.03**	0.99	0.43	0.07
SFOU	5	0.24	0.92	0.74	**0.01**	**-0.01**	**0.04**	**-0.01**	1.00	0.16	0.03

[1]LV = optimal number of latent variables selected; RMSECV = root mean squared error cross-validation (%); R^2_p = prediction sum of squares (PRESS R^2); RMSEP = root mean squared error prediction for n = 56 (%); RSR = RMSE-observation standard deviation ratio (optimal RSR < 0.5); .

[2]AFMP = adulterated fat-filled milk powder; COM, COU, POM, POU, SBOM, SBOU, SFOM and SFOU are the FMP types: coconut oil (CO) with melamine, CO with urea, palm oil (PO) with melamine, PO with urea, soya-bean oil (SBO) with melamine, SBO with urea, sunflower oil (SFO) with melamine and SFO with urea.

[3]SG + EMSC Full = Savitzky-Golay (second derivative with second polynomial order and a window size of 25 nm) plus the extended multiplicative signal correction pre-processing methods performed on the full spectral range, 850 – 2500 nm of the calibration sets (n = 184).

[4]SG + EMSC iPLS = Savitzky-Golay (second derivative with second polynomial order and a window size of 25 nm) plus the extended multiplicative signal correction pre-processing methods performed only on the selected sensitive wavelength intervals of the calibration sets using the interval PLS variable selection. [5]Bold face represents the predictions with the lowest relative errors (< 0.20) at 6 concentration levels.

5.4 Conclusion

The similarity in profiles between the typical and the atypical PO FMP types was established through the PCA and the mean spectra analyses. Both analyses also established the dissimilarity in profiles between the adulterated and the unadulterated CO FMP types, thus highlighting the contrasting abilities of palm and coconut oils in masking the presence of melamine and urea adulterations in FMP. The multilevel SIMCA models built using HMS were able to detect, confirm and differentiate the adulterations with an excellent efficiency. In particular, the one-class SIMCA model for CO FMP type was able to screen the studied chemical adulterants with maximum efficacy (i.e., sensitivity and specificity values of 100 %). Consequently, this study has demonstrated that NIRS is potentially useful for effective detection of melamine and urea in FMP especially in those formulated with coconut oil where the LOD was found to be 0.01 %. However, further studies are recommended to assess the efficacy of NIRS for the detection of these adulterants at lower concentration levels. On the other hand, the PLSR models developed for the quantification of adulterants in FMP yielded satisfactory predictions ($R^2_p \geq 0.96$ and RSR ≤ 0.19) at 1.00 % adulteration level. Overall, melamine adulterations were better detected and quantified than that of urea in the FMP samples. Also, the variable selection techniques substantially improved the robustness of the pre-processed models yielding lower errors and better predictions. Our results suggest that cost-efficient screening of FMP products for adulterants can be achieved with the approach described herein, and hope their wider adoption helps to protect public health, notably in countries where such products are an important protein source but also where regulatory networks may have limited resources.

Chapter 6

Paper IV

Characterisation of the quality alterations in model fat-filled milk powders under inclement conditions and the prediction of the storage time using near infrared spectroscopy

Abstract

Fat-filled milk powders (FMP) are exported to tropical developing markets as inexpensive milk alternatives. Consequently, FMP are exposed to high temperature and humidity over long distribution and storage times, presenting challenges in preserving product quality and stability. Efficient and cost-effective methods for quality assurance under such conditions are needed. We utilised the changes in profile of the fatty acids, amino acids and near infrared spectra to investigate the quality alterations in 4 types of FMP produced onsite with 4 different vegetable oils (i.e., coconut, palm, soya-bean and sunflower) and stored for 7 weeks at 40°C. Stearic acid decreased while the leucine content increased upon storage, but palm oil FMP appeared to be the most stable. Multiclass analyses offered substantive separation between the fresh/aged samples. The models based on interval-PLS efficiently (NSE ≥ 0.90) predicted storage time with low errors (RSR ≤ 0.28), indicative of FMP freshness and stability.

6.1 Introduction

Fat-filled milk power (FMP) is a recombined product resulting from the blending of skimmed milk powder (SMP) with vegetable fat (Codex Alimentarius Commission, 2006). It is formulated in temperate countries and exported in bulk to developing markets including many African nations where they are repacked into single serve/small multilayer laminated pouches (European Commission, 2017) for storage and retail distribution. Consequently, the repackaged FMP, with high fat and protein contents (≥ 26 % (w/w) and ≥ 34 % (w/w) respectively (Codex Alimentarius Commission, 2006)), is exposed to harsh climatic conditions (i.e., high temperature and humidity) over a long distribution and storage times, making it difficult to preserve product quality and stability, especially in tropical countries where temperature and humidity are relatively high (Uppu, 2001). The storage behaviour of FMP may strongly depend on the protein composition and the degree of unsaturation of the incorporated vegetable oils. However, it has long been known (Ford et al., 1983) that the storage of milk powder under unfavourable conditions accelerates the normally slow deterioration in nutritional quality. Some of the most important deteriorative mechanisms in milk powder during storage are lactose crystallisation, non-enzymatic browning (i.e., Maillard reaction) and lipid oxidation; and the consequences of these processes include particle collapse and caking, loss of flavour and essential nutrients, increased protein interactions and aggregations, decreased freshness and the generation of oxidative products (Thomas et al., 2004). For crude and refined vegetable oils, autoxidation or oxidative rancidity is known to be the major cause of quality losses during storage (Crapiste, Brevedan, & Carelli, 1999). However, proteins have been found (Estévez, Kylli, Puolanne, Kivikari, & Heinonen, 2008) to inhibit lipid oxidation in oil-in-water emulsions and the degree of their antioxidant/prooxidant effects depends on their amino acid (AA) composition. To the best of our knowledge, the changes in quality that may occur in FMP under environmental stress, in the midst of the interactions between the proteins and the vegetable oil lipids, have never been reported. In addition, as FMP is subjected to different climatic conditions and stresses along the distribution chain (Uppu, 2001), there is a great need to independently predict the length of time for which it has been in storage, to monitor product shelf-life and quality, notably in the developing markets where border controls are poor (FAO/WHO, 2003), and milk supplies involve many intermediaries.

The potential for product deterioration in undesirable conditions is a strong driver for a rapid, inexpensive method for assessing the shelf stability of FMP during storage. Al-Qadiri et al. (2008) successfully used near infrared spectroscopy (NIRS) in combination with chemometric methods to rapidly detect and monitor the spoilage of pasteurised skim milk during storage; however, NIRS has not to our knowledge been used for monitoring the behaviour of FMP over time under the conditions of environmental stress. The applicability of NIRS might be challenged by the many physicochemical damages that occur in milk powder matrices during storage (Thomas et al., 2004).

In this study, NIRS coupled with chemometric techniques were used to characterise the quality decay in FMP produced onsite and stored under inclement conditions; and to predict the storage time, indicative of the milk powder freshness. The changes in the contents of the fatty acid (FA) and the AA profiles of FMP were also evaluated to understand the influence of the type of incorporated vegetable oil on the milk powder stability during prolonged storage.

6.2 Material and methods

6.2.1 Raw materials collection

The four vegetable oils selected for this study contained low (e.g., coconut oil (ca. 8 %)), medium (e.g., palm oil (ca. 44 %)) and high (e.g., soya-bean oil (ca. 84 %) and sunflower oil (ca. 90 %) concentrations of unsaturated fatty acids (UFA). Palm oil (PO) was purchased from Pure Nature (New Zealand) whereas coconut oil (CO), soya-bean oil (SBO), sunflower oil (SFO) and SMP were purchased from a local supermarket.

6.2.2 Model fat-filled milk powder manufacture, packaging and storage

A batch of FMP was freshly formulated for each of the four FMP types (i.e., CO, PO, SBO and SFO) using the methods described previously (Ejeahalaka & On, 2019a). The moisture content was determined as 4.1 ± 0.25 g per 100 g according to ISO 5537-2004 protocol. Baseline NIRS readings were established (see Section 2.4) using 20 aliquots of each freshly manufactured FMP type. Furthermore, 21 aliquots (each 10 g) were immediately taken from each of the four fresh FMP types and packed/sealed individually in single serve flexible laminate pouches (Pack-fresh Ltd, New Zealand) under normal atmospheric conditions, thus resulting in 84 samples for the storage stability trial. The barrier properties of the laminate pouches were given as: water vapour transmission rate (2.20 g m^{-2} day^{-1} measured at 38 °C and 90 % relative

humidity (RH) according to ASTM F1249) and oxygen transmission rate (1.40 cc m^{-2} day^{-1} bar^{-1} measured at 23 °C and 0 % RH according to ASTM F2622). The packaging material for this study was utilised under small-scale laboratory conditions and likely differ from those used by large-scale commercial manufacturers.

The sealed pouches containing the aliquots of the fresh FMP samples were carefully inspected to ensure that there were no leakages before they were stored away from direct light in a Contherm incubator (Contherm Scientific Ltd, New Zealand) at a temperature of 40 ± 1 °C and RH of 90 ± 5 % for a period of 7 weeks. The milk powder samples were placed on a dry plastic tray, positioned above a temperature-controlled reverse-osmosis purified water-bath, within a large sealable plastic container with lid, in order to create the saturated environment. The temperature and RH data were recorded in real time using sensors and loggers. High target values of temperature (40 °C) and RH (90 %) were chosen for the trial to replicate the tropical conditions and to accelerate the deteriorative changes in FMP within the shortest possible time. The stored/aged FMP samples containing coconut oil, palm oil, soya-bean oil and sunflower oil were denoted as CO-A, PO-A, SBO-A and SFO-A respectively for ease of reference. Sampling involving NIRS measurements was done weekly in triplicates and each of the pouches was measured once. Chemical analyses (i.e., FA and AA profiling) were also performed in triplicates at the beginning of the trial (t = 0 week) on fresh samples and at the end of the 7 weeks of storage.

6.2.3 Fat-filled milk powder chemical analyses

6.2.3.1 Fatty acid determination

The methods employed for the determination of FA concentrations have been described in detail previously (Ejeahalaka & On, 2019a). In brief, a one-step methylation process for 0.15 g aliquots of the freeze-dried FMP samples using 900 µL of heptane and 4.0 mL of 0.5 M NaOH/dried methanol in Kimax tubes, followed by further addition of 2.0 mL of heptane and 2.0 mL of distilled water and subsequent use of small amounts of anhydrous sodium sulphate to remove residual water, was used. Subsamples of extracts containing FA methyl esters (FAME) were stored in vials at - 20 °C until GC analysis.

The GC analysis was carried out by injecting 1.0 µL of the extracted FAME onto a Varian CP7420, tailor-made fused silica capillary column with a length of 100 m, internal diameter of 0.25 mm, and film thickness of 0.20 µm, using the AOC-20i auto-sampler fitted to a Shimadzu

GC-2010 gas chromatograph, as described previously (Ejeahalaka & On, 2019a). The FA concentrations obtained from the assay were expressed as g of fatty acid per 100 g of total fatty acids. However, only the FA of the aged FMP samples were measured in the present study as those for the Fresh FMP types had been previously reported (Ejeahalaka & On, 2019a). The changes in the FA concentrations and the lipid indices of atherogenicity and thrombogenicity (Ulbricht & Southgate, 1991) were estimated to aid understanding of the phase changes and the oxidation processes that may have occurred as a result of the incorporated vegetable oils in FMP during prolonged storage under inclement condition.

6.2.3.2 Amino acid profiling

The profiling of the AA composition in the samples was done in two steps: acid hydrolysis and high-performance liquid chromatograph (HPLC) analysis. The acid hydrolysis was performed by weighing 0.1 g of the freeze-dried milk samples directly into the culture tubes in triplicates. Then, 10.0 µL 0.5 M of internal standard (amino-butyric acid) was added into each of the tubes, followed by the addition of 5.0 mL 6 N of hydrochloric acid. The tubes, with their caps screwed, were carefully vortexed and allowed to sit in ultrasonic bath for 5 minutes. Then, each tube was purged with nitrogen (oxygen free) for 30 seconds and immediately screwed tightly with Teflon cap before putting in block heater (Ratek, Australia) and heating at 110 °C for 20 hours. The tubes were allowed to cool down to room temperature before they were loaded onto the EZ-2 vacuum centrifuge (GeneVac Scientific, USA) at a maximum temperature of 47 °C. When the drying process was completed, the dried residue was reconstituted and rinsed with nano-pure water into a 50ml volumetric flask. The hydrolysate solution was then filtered through a 0.45 µm syringe filter into a 2 mL vial and kept frozen at – 20 ºC in triplicates until HPLC analysis.

The HPLC analysis was carried out by Agilent 1100 series HPLC system with EZChrom Elite software (Agilent Technologies, Walbronn, Germany), equipped with a binary pump, and an auto-sampler with thermostat. 11.0 µL of pre-column derivatised sample was injected onto an ACE C-18 column (ACE, UK) of dimension 150 x 4.6 mm and particle size 3 µm. The column temperature was kept at 40 °C. The mobile phase A consisted of 0.01 M Na_2HPO_4 with 0.8 % THF adjusted to pH of 7.5 with H_3PO_4, whereas the mobile phase B consisted of 50 % methanol and 50% acetonitrile, flowrate 0.7 mL/min. The pre-column derivatisation was performed on the autosampler using: o-phthaldialdehyde for primary amino acid, and 9-fluorenylmethyl chloroformate for secondary amino acid. The derivatised amino acids were detected by

fluorescence detector, set at excitation and emission wavelengths of 335 nm and 440 nm respectively. Then, the detector was switched to excitation and emission wavelengths of 260 nm and 315 nm respectively at 21 min in order to detect proline. The separation was completed in 36 min. The standard curve was graphed by using individual amino acids standards. The results obtained were expressed in g of amino acids per 100 g of freeze-dried milk samples. The AA concentrations of the fresh and aged samples were measured in order to understand the changes in protein that occurred as a result of the strong interactions between the vegetable oil lipids and the proteins in the FMP types under inclement condition.

6.2.4 Near infrared spectral collection

The near infrared spectra of the fresh FMP (i.e., CO, PO, SBO and SFO) and that of the FMP types stored/aged for a period of 7 weeks (i.e., CO-A, PO-A, SBO-A and SFO-A) were measured immediately using a FOSS NIRSystem (model DS 2500F) spectrometer as described previously (Ejeahalaka & On, 2019a). Briefly, 5 g of each of the milk samples was placed in a 10-mL DS 2500 ring cup (cup type: 2004) and scanned at a resolution of 0.5 nm, wavelength accuracy of < 0.05 nm, analysis time of < 1 min using a detector array that was composed of both silicon (850 – 1100 nm) and lead sulfide (1100 – 2500 nm). A total of 164 unique objects (comprising 20 samples from each of the 4 fresh FMP types; and 21 from each of the 4 stored/aged FMP types) were scanned consecutively in triplicates using the same ring cup without refilling at each scan to minimise sampling error and enhance repeatability. Each spectrum acquired comprised of 3300 absorbance values recorded from 850 to 2499.5 nm averaging 32 scans. The triplicate spectra collected from each sample were averaged and then subjected to chemometric analysis. The coefficients of variation were computed across the 3 replicate spectra as previously described (Ejeahalaka & On, 2019a).

6.2.5 Chemometric and statistical analysis

The raw NIRS scans of the 164 unique samples were transferred into the R software version 3.5.1 (R Core Team, 2018) and they were immediately pre-processed with the extended multiplicative signal correction (EMSC) (Stark & Martens, 1996) technique before plotting the mean spectra (i.e., the mean of the absorbance values across the wavelengths for all the samples) to highlight the subtle changes in profiles of the 4 FMP types across the 7 weeks of storage. The raw NIRS scans were also subjected to principal component analysis (PCA) after mean-centring to identify any natural groupings within the spectral data. The major effect of

mean-centring was to remove the broad sloping background from the data in order to improve the interpretability of the model (Boysworth & Booksh, 2001). It was also meant to project the first principal component onto the direction of greatest variance within the spectra data (Boysworth & Booksh, 2007). Furthermore, the PCA loadings was plotted to identify the optimal wavelengths that carried most of the information in the spectra data (J. P. Wold, Jakobsen, & Krane, 1996). To build the classification rules for differentiating between the fresh/aged FMP types and to predict the storage time, the raw spectra were pre-processed using a second derivative transformation with a Savitzky-Golay (SG) algorithm (window size = 25 points, second-order polynomial fit) (Savitzky & Golay, 1964) followed by EMSC in order to suppress the broad underlying baselines and to remove the multiplicative effects. Zimmermann and Kohler (2013) demonstrated that the use of SG with EMSC for pre-processing results in simpler and often better models, since the SG differentiation effectively suppresses the broad underlying baselines, while the EMSC principally has the feature of removing the multiplicative effect. Several researchers (Bruun et al., 2007; Ejeahalaka & On, 2019b; Vongsvivut et al., 2012) have regularly used the SG differentiation with EMSC pre-processing technique to reduce the prediction errors and improve the results. Thereafter, the pre-processed spectra were divided into a training set (for building the models) and a testing set (for independent valuation of the models) in the ratio of 4:1 (Ejeahalaka & On, 2019a) using the random selection (RS) method and the Kennard-Stone (KS) algorithm. The RS method employs a simple random splitting process whereas the KS algorithm technique selects samples sequentially based on their Euclidian distances (Kennard & Stone, 1969).

Statistical analyses were performed by paired *t*-test to determine whether there was a change in the FA and AA contents of the four FMP types formulated at t = 0 week (4 °C) and stored at 40 °C for t = 7 weeks (significance level $p < 0.05$). PCA analyses were also conducted to evaluate the influence of storage on the FA and the AA contents of the FMP samples.

6.2.5.1 Aged fat-filled milk powders characterisation

Soft independent modelling of class analogy (SIMCA) was used to predict the class memberships and to describe how much the aged samples resembled or differed from themselves and from the fresh FMP types. Consequently, five one-class SIMCA models were built using the SG + EMSC pre-processed spectra for each of the following four different treatments: a) RS sampling method with the full spectral range (850 – 2500 nm, 3300 wavelengths); b) KS algorithm sampling technique with the full spectral range; c) RS sampling

method based only on the sensitive wavelengths selected from the PCA loadings; and d) KS algorithm sampling technique based only on the sensitive wavelengths selected from the PCA loadings. The SIMCA models established a PCA for each of the classes (S. Wold, 1976) and the optimal number of principal components (PC) was determined through internal validation using the leave-one out cross validation method (Ejeahalaka & On, 2019a). The models were built with 124 samples (i.e., 15 for each of the 4 fresh FMP types and 16 for each of the 4 aged FMP types) for calibration/internal validation and 40 samples (i.e., 5 for each of the 4 fresh FMP types and 5 for each of the 4 aged FMP types) for independent testing in order to evaluate their performances. The values of sensitivity (i.e., the rate at which the modelled class recognises its members), specificity (i.e., the rate at which the modelled class rejects samples not belonging to it), efficiency (i.e., the geometric mean of the values of sensitivity and specificity) (Chong & Jun, 2005) and the reliability rate (i.e., the summation of the values of sensitivity and specificity minus 100 %) (Reis et al., 2017) were used to assess model performances. In particular, the efficiency (E) and the reliability rate (RR) provided the figures of merit of the models. These metrics were used for similar assessment in our previous study (Ejeahalaka & On, 2019b).

6.2.5.2 Quantitative prediction of the storage time

Partial least square regression (PLSR) was used to correlate the absorbance values of the aged FMP samples (i.e., CO-A, PO-A, SBO-A and SFO-A) to their corresponding storage time (in weeks) across the 3300 wavelengths, in accordance with Lambert-Beer's law. PLSR model was built for each of the four aged FMP types using the SG + EMSC pre-processed spectra with a combination of either the full spectral range (850 – 2500 nm, 3300 wavelengths) and the RS sampling method / the KS algorithm technique or the dominant spectral interval(s) obtained from the interval partial least square (iPLS) variable selection (Nørgaard et al., 2000) and the RS sampling method / the KS algorithm technique as previously described for the SIMCA models above (Section 6.2.5.1). The purpose of using the iPLS technique was to find one or a few intervals which give better predictions than those obtained when using the full spectrum, and to provide an overview to help in spectroscopic interpretation (Andersen & Bro, 2010). To build the PLSR models, a total of 64 samples (i.e., 16 for each of the 4 aged FMP types) were used as the calibration/internal validation set and 20 samples (i.e., 5 for each of the 4 aged FMP types) as an independent test set. Leave-one out cross validation method was used for internal validation to determine the optimal number of latent variables (LV) required to build

the models. The prediction capacity of the calibrations was assessed using the root mean squared error of cross validation (RMSECV). The relevant statistics used for computing the performances of the models on independent test set include: coefficient of determination of prediction (R^2p); root mean squared error of prediction (RMSEP); percent bias (PBIAS; i.e., the average deviation of the predicted data from the measured values)(Moriasi et al., 2007); RMSE-observation standard deviation ratio (RSR) and the Nash-Sutcliffe efficiency (NSE). However, the comparison of the overall performances of the models were based on RSR and NSE. RSR was successfully used in our previous study (Ejeahalaka & On, 2019b) for similar assessment.

RSR is the ratio of the root mean squared error and standard deviation of measured data (Moriasi et al., 2007). Hence,

$$RSR = \frac{RMSEP}{STDEV} = \sqrt{\frac{(n-1)}{n} \times \frac{\sum_{i=1}^{n}(P_i - M_i)^2}{\sum_{i=1}^{n}(M_i - M)^2}} \quad (6.1)$$

where M_i is the *i*th measured value, P_i the *i*th predicted value and M the measured mean values for the *i*th sample. RSR provides a delimiter of what is considered a low RMSEP based on the standard deviation of the *n* measured values (Moriasi et al., 2007). It varies from an optimal value of 0, which indicates zero RMSEP and therefore perfect model prediction, to a large positive value (Moriasi et al., 2007)

On the other hand, NSE determines the relative magnitude of the residual variance ("noise") compared to the measured data variance ("information") and as such reflects the overall goodness of fit of the model (Nash & Sutcliffe, 1970).

$$NSE = 1 - \left[\frac{\sum_{i=1}^{n}(M_i - P_i)^2}{\sum_{i=1}^{n}(M_i - M)^2}\right] \quad (6.2)$$

NSE ranges from -∞ to 1 with NSE = 1 being the optimal value (Moriasi et al., 2007).

6.3 Results and discussion

6.3.1 Changes in FA and AA profiles of FMP types under environmental stress

The behavioural differences between the FMP types subjected to high temperature (i.e., 40 ± 1 °C) and relative humidity (i.e., 90 ± 5 %) for 7 weeks were characterized by FA and AA

analyses, and the results obtained are shown in Table 6.1 and Table 6.2 respectively. As can be seen from Table 6.1, the total saturated fatty acids (SFA) content of the CO FMP type increased by 0.93 % with caprylic acid (C8:0) and stearic acid (C18:0) recording the highest increments; whereas those for PO, SBO and SFO decreased by 0.41 %, 3.82 % and 4.84 % respectively with C18:0 as the most dominant. Thus, the adverse storage conditions reduced the C18:0 content of FMP for each of the milk fat replacements with vegetable oils and this reduction was found to be statistically significant ($t(3) = -9.70$, $p = 0.002$) according to the paired *t*-test result in Table B1 (Appendix B). On the other hand, the total unsaturated fatty acids (UFA) content of CO decreased by 2.80 % whereas those for PO, SBO and SFO increased by 0.95 %, 1.13 % and 1.10 % respectively. Thus, the CO FMP type had a decreased UFA and an increased SFA, in stark contrast to the results obtained for PO, SBO and SFO. In addition, the index of atherogenicity (IA) for CO FMP type increased by 3.59 % while that for PO decreased by 1.67 %. No changes in IA was recorded for SBO and SFO FMP types under the same storage conditions, although a greater change was found in their linolenic acid than their linoleic acid. Overall, the PO FMP type appeared to be the most stable as it had the least changes in FA. However, the replacement of milk fat with palm oil (or the other vegetable oils studied) did not make the product more stable as whole milk powder (WMP) recorded slightly lesser changes than the PO FMP type under the same environmental stress. WMP followed the same profile pattern (i.e., decreased SFA and increased UFA) as PO, SBO and SFO FMP types upon storage under the same conditions. However, the decrease in oleic acid and linoleic acid that was found for CO FMP type was similar in pattern to that of powdered infant formula reported by (Rodríguez-Alcalá et al., 2007). The authors found that there was a significant reduction in oleic and linolenic acids during the first year of storage of the milk powder. Overall, the paired *t*-test result (Table B1) showed that C18:0 was the only FA content that recorded a statistically significant change over the 7 weeks of storage of the FMP types.

The results presented in Table 6.2 and Table B2 (Appendix B) showed that all the FMP types recorded a statistically significant increase in leucine content ($t(3) = 6.39$, $p = 0.008$), an observation analogous to that seen previously in an investigation of the stability of whole milk powder produced from raw milk reverse osmosis retentate (Sørensen et al., 2017). We also saw a statistically significant decrease in most of the AA (i.e., lysine ($t(3) = -16.47$, $p = 0.000$), tryptophan ($t(3) = -4.24$, $p = 0.024$), isoleucine ($t(3) = -15.21$, $p = 0.001$), methionine ($t(3) = -5.67$, $p = 0.011$), phenylalanine ($t(3) = -5.44$, $p = 0.012$), glycine ($t(3) = -4.82$, $p = 0.017$), alanine

(t(3) = -5.62, p = 0.011), aspartate (t(3) = -10.22, p = 0.002), serine (t(3) = -5.46, p = 0.012) and tyrosine (t(3) = -5.17, p = 0.014)) contents after the 7 weeks of storage. In addition, the FMP types had a statistically significant decrease in essential AA (t(3) = -7.52, p = 0.005) and non-essential AA (t(3) = -3.63, p = 0.036). A close inspection revealed that the percentage change in both the essential and the conditionally essential AA concentrations decreased with increasing unsaturation of the incorporated vegetable oils in FMP. Thus, the incorporated sunflower oil had the largest protective influence on the essential AA contents of FMP while coconut oil had the least. On the other hand, WMP, containing milk fat, incurred lower losses in lysine, threonine, alanine, aspartate and serine contents (data not shown) than FMP containing vegetable oils (i.e., CO, PO, SBO and SFO) under the same condition.

Figs. B3 and B4 (Appendix B) show the PCA score plots and biplots of the FA and AA profiles of the fresh/aged FMP types respectively. As can be seen from the figures, the profiles of the FMP types before (at t = 0 week) and after (at t = 7 weeks) storage were clearly distinguishable. In particular, the fresh and the aged FMP types showed greater separation in the plots of their AA profiles (Fig. B4) than in those of their FA concentrations (Fig. B3). Thus, the inclement storage conditions (i.e., 40 °C, 90 % RH) probably had greater influence on the AA than on the FA profiles of the FMP samples. The biplot of the AA profiles (Fig. B4(b)) shows that serine was associated with CO-A, cysteine with PO-A, leucine with SBO-A, and histidine/glutamic acid with SFO-A FMP types. On the other hand, the first principal component (PC 1) was negatively associated with serine, cysteine and leucine. For the FA profiles, the biplot (Fig. B3(b)) shows that C16:0, C8:0, C10:0, C12:0 and C14:0 were associated with CO and CO-A; C16 and C18:0 with PO; C18:2 with SBO and SBO-A; and C18:3 with SFO and SFO-A FMP types.

Table **6.1:** Changes (%) in fatty acid (FA) concentrations (expressed as g/100 g of FA) of fresh[1] fat-filled milk powers (FMP) types[2] formulated from 4 different vegetable oils and stored for 7 weeks at 40 °C and 90 % relative humidity

		Concentrations after 7 weeks of storage							
		CO-A		PO-A		SBO-A		SFO-A	
FA	WMP[3]	Value[4]	Change	Value	Change	Value	Change	Value	Change
Saturated Fatty Acids (SFA)									
C6:0	-13.16	0.43	-4.08	0.00	-	-	-	-	-
C8:0	-7.89	7.23	7.61	0.01	-	-	-	-	-
C10:0	-4.51	5.11	1.31	0.02	-	-	-	-	-
C12:0	-2.73	48.38	2.09	0.20	-17.91	-	-	-	-
C14:0	-0.83	19.72	0.59	0.71	-16.55	0.07	-13.55	0.07	-14.00
C16:0	0.99	9.14	-5.43	49.27	0.74	10.53	-1.79	5.84	-2.69
C18:0	2.70	2.66	-9.64	4.64	-8.61	3.39	-9.65	2.75	-9.66
∑SFA	-0.24	92.67	0.93	54.84	-0.41	13.99	-3.82	8.68	-4.84
Unsaturated Fatty Acids (UFA)									
OA	1.59	6.01	-2.77	32.98	2.18	21.91	0.38	30.76	2.24
LA	-1.24	1.67	-2.94	10.83	-2.61	55.15	0.23	57.79	0.51
ALA	-1.56	-	-	0.21	0.75	5.82	13.90	0.25	-1.52
∑UFA	1.30	7.68	-2.80	44.02	0.95	82.89	1.13	88.80	1.10
Ratios and indices[5]									
MUFA	1.59	6.01	-2.77	32.98	2.18	21.91	0.38	30.76	2.24
PUFA	-1.39	1.67	-2.94	11.05	-2.55	60.97	1.40	58.04	0.50
OA/LA	2.87	3.60	0.18	3.04	4.92	0.40	0.15	0.53	1.72
IA	-1.54	17.75	3.59	1.18	-1.67	0.13	0.00	0.07	0.00
IT	0.18	8.20	0.67	2.41	-1.23	0.25	-7.41	0.19	-5.00

[1]Details of the FA concentrations of the fresh FMP types at t = 0 week can be found at Ejeahalaka & On (2019)

[2]CO-A, PO-A, SBO-A and SBO-A represent the coconut oil, palm oil, soya-bean oil and sunflower oil FMP types that were aged/stored at 40 °C for t = 7 weeks.

[3]Change recorded in whole milk powder (WMP) aged/stored at 40 °C for t = 7 weeks

[4]Each fatty acid (FA) concentration value represents the mean of triplicate measurements (n = 3).

[5]MUFA = monounsaturated FA, PUFA = polyunsaturated FA, OA = oleic acid, LA = linoleic acid and ALA = alpha-linolenic acid. Indices of atherogenicity (IA) and thrombogenicity (IT) were estimated according to Ulbricht & Southgate (1991).

Table **6.2:** Changes (%) in amino acid (AA) concentrations (expressed as g/100 g of milk) of fresh fat-filled milk powers (FMP) types[1] formulated from 4 different vegetable oils and stored for 7 weeks at 40 °C and 90 % relative humidity

		Concentrations after 7 weeks of storage							
		CO-A		PO-A		SBO-A		SFO-A	
AA	Fresh[2]	Value[3]	Change	Value	Change	Value	Change	Value	Change
Essential Amino Acids (EAA)									
Lys	3.36	2.82	-16.07	2.89	-13.99	2.88	-14.29	2.96	-11.90
Thr	1.38	1.26	-8.70	1.15	-16.67	1.31	-5.07	1.32	-4.35
Trp	0.26	0.21	-19.23	0.24	-7.69	0.23	-11.54	0.24	-7.69
Leu	1.57	1.69	7.64	1.84	17.20	1.78	13.38	1.76	12.10
Ile	3.14	2.64	-15.92	2.74	-12.74	2.70	-14.01	2.77	-11.78
Val	1.71	1.62	-5.26	1.70	-0.58	1.66	-2.92	1.68	-1.75
Met	0.74	0.64	-13.51	0.62	-16.22	0.68	-8.11	0.68	-8.11
Phe	1.50	1.36	-9.33	1.44	-4.00	1.41	-6.00	1.42	-5.33
His	0.93	0.90	-3.23	0.93	0.00	0.99	6.45	0.97	4.30
∑EAA	14.59	13.14	-9.94	13.55	-7.13	13.64	-6.51	13.80	-5.41
Conditionally Essential Amino Acids (CEAA)									
Arg	1.21	0.77	-36.36	0.81	-33.06	1.11	-8.26	1.11	-8.26
Pro	3.38	3.22	-4.73	3.38	0.00	3.28	-2.96	3.35	-0.89
Cys	0.17	0.17	0.00	0.19	11.76	0.17	0.00	0.18	5.88
Gly	0.64	0.52	-18.75	0.58	-9.38	0.57	-10.94	0.59	-7.81
Tau	0.02	0.02	0.00	0.02	0.00	0.02	0.00	0.02	0.00
∑CEAA	5.42	4.70	-13.28	4.98	-8.12	5.15	-4.98	5.25	-3.14
Non-essential Amino Acids (NEAA)									
Ala	1.15	1.00	-13.04	1.08	-6.09	1.05	-8.70	1.07	-6.96
Asp	2.23	1.98	-11.21	2.07	-7.17	2.02	-9.42	2.05	-8.07
Glu	5.77	5.63	-2.43	5.89	2.08	5.89	2.08	5.80	0.52
Ser	1.90	1.70	-10.53	1.80	-5.26	1.78	-6.32	1.80	-5.26
Tyr	1.38	1.29	-6.52	1.35	-2.17	1.31	-5.07	1.31	-5.07
∑NEAA	12.43	11.60	-6.68	12.19	-1.93	12.05	-3.06	12.03	-3.22

[1]CO-A, PO-A, SBO-A and SBO-A represent the coconut oil, palm oil, soya-bean oil and sunflower oil FMP types that were aged/stored at 40 °C for t = 7 weeks.

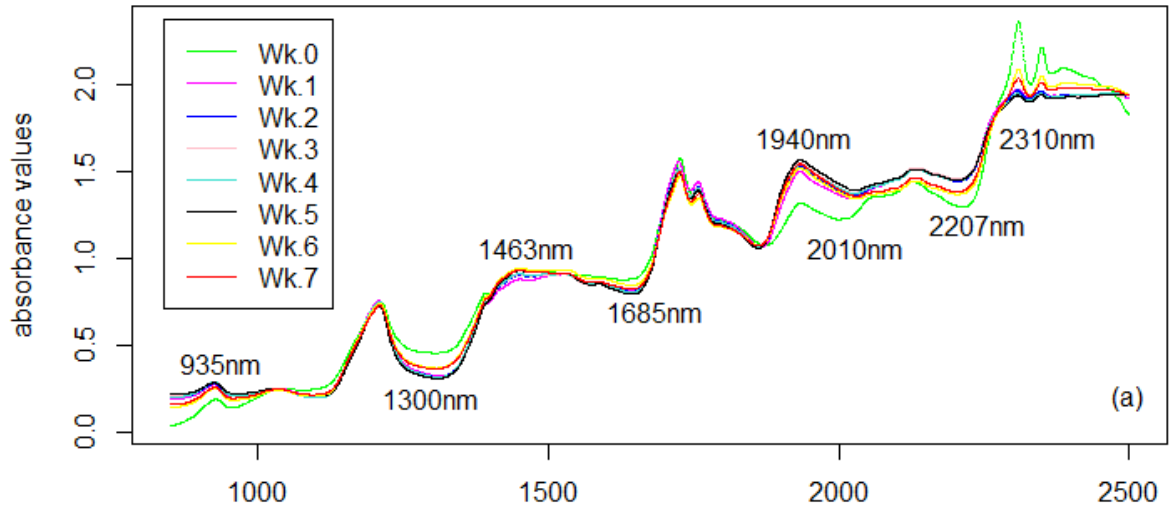
[2]Fresh refers to the FMP types prior to aging at t = 0 week with each of the samples having the same amino acid concentrations regardless of type.

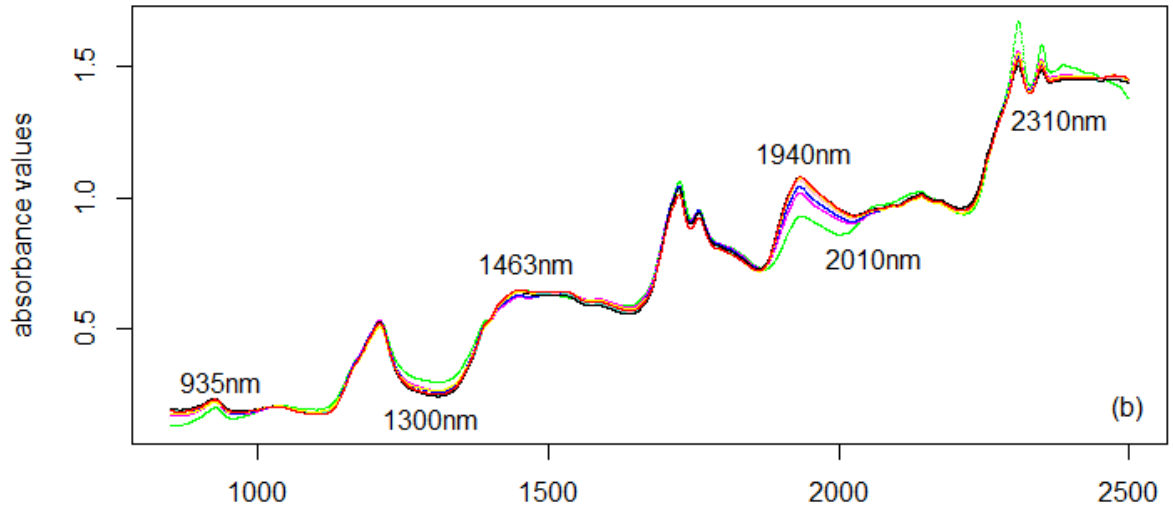
[3]Each amino acid concentration value represents the mean of triplicate measurements (n = 3).

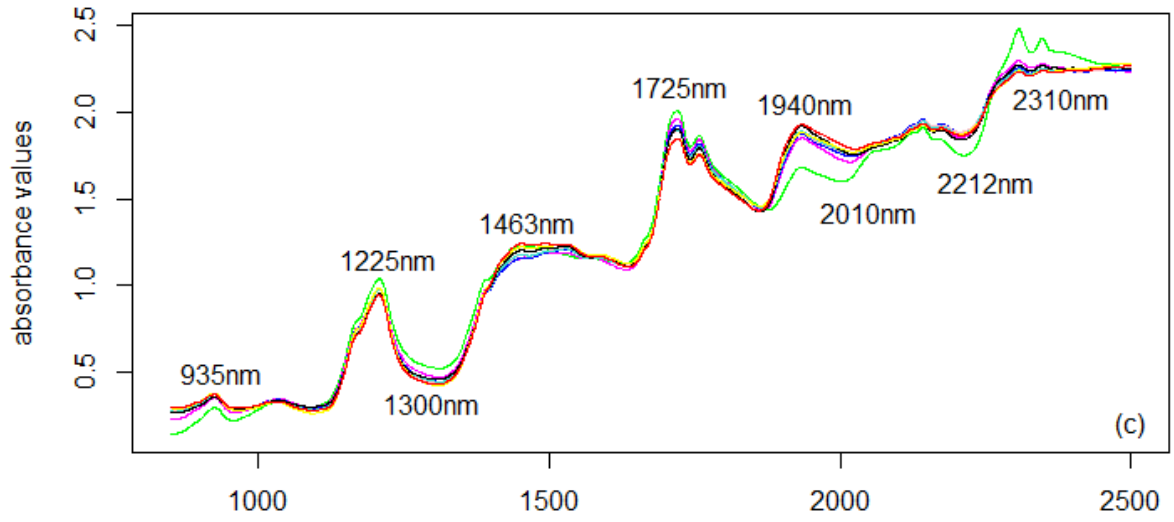
6.3.2 NIRS profile alterations in response to FMP accelerated storage

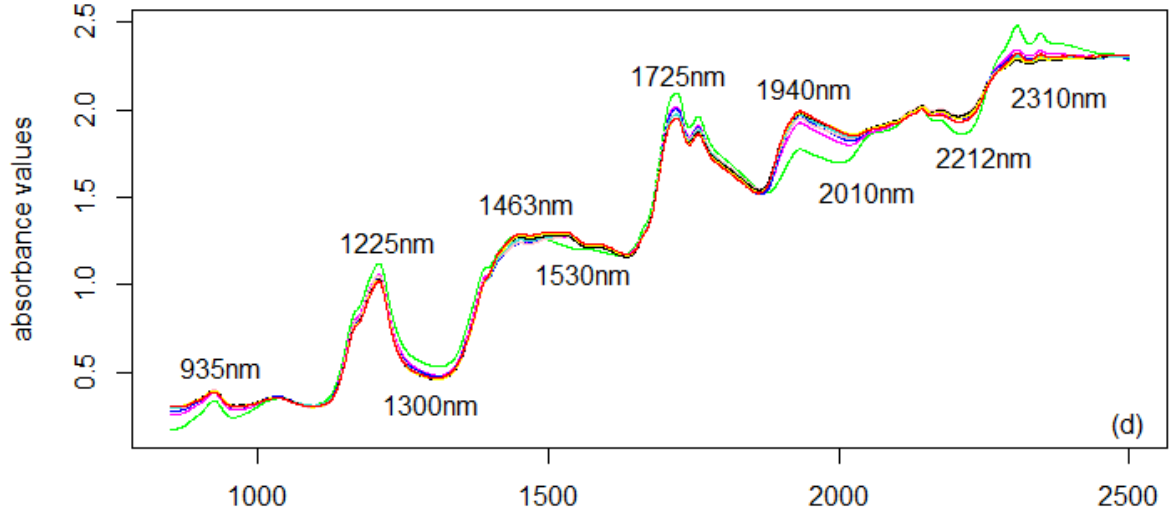
The coefficients of variation across the wavelengths of the triplicate spectra for each of FMP samples were < 4.0 %. Figure 6.1 shows the EMSC pre-processed near-infrared spectra of the FMP types stored for 7 weeks at 40° C and 90 % RH. The fresh (green spectrum, t = week 0) and aged (t = week 1-7) samples yielded similar patterns, with occasional shifts in absorption peaks, emphasising the utility of NIRS as a quality assurance method to monitor changes in FMP during storage under inclement conditions. Some of the wavelengths at which significant alterations occurred were as indicated on the spectra profiles. As can be seen from the figure,

the absorbance bands increased with storage time at about 935 nm, 1530 nm and 1940 nm; and decreased at about 1225 nm, 1300 nm, 1685 nm, 1725 nm, 2010 nm, 2207 nm, 2212 nm and 2310 nm. According to Workman and Weyer (2012), these wavelengths may be associated with the following: 1225 nm due to second overtone of an aliphatic methane (under C-H group), 1463 nm due to first overtone of N-H (under N-H group), 1530 nm due to first overtone of N-H stretching (under N-H group), 1685 nm due to first overtone of aromatic C-H absorptions (under C-H group), 1725 nm due to first overtone of C-H stretching of methylene (under C-H group), 1940 nm due to combination of the asymmetric stretch and bending of the water molecule (under O-H group), 2010 nm due to combination of symmetric NH stretching with amide III (under N-H group), 2207 nm due to peptide N-H and C=O groups at right angles to the line of the peptide backbone, 2212 nm due to C=O/C-N/N-H combination, and 2310 nm due to C-H bending typical of lipids (under C-H group). Thus, the peak intensity of the O-H functional group increased with storage time; while that of the C-H and C=O functional groups decreased with storage time. It appears that CO FMP type had unique band alterations near 1685 nm and 2207 nm. Similarly, the SFO FMP type had a relatively definable characteristic band alteration (i.e., a slight decrease in absorption intensity) at about 1530 nm. However, the alterations near 1225 nm, 1725 nm and 2212 nm were peculiar to SBO and SFO FMP types and as such may be useful in discriminating the aged samples containing the saturated and the unsaturated vegetable oils. Further examination revealed that there was a slight but steady increase in the absorption intensities of the aged CO, SBO and SFO FMP types from about 2010 nm to 2212 nm and those alterations included the band near 2070 nm which was previously (Holman, Nickell, Privett, & Edmondson, 1958; Wójcicki, Khmelinskii, Sikorski, & Sikorska, 2015) attributed to the formation of hydroperoxides, indicative of oxidative stress, in edible oils. In general, the spectra of PO FMP type appeared the most stable across the 7 weeks of storage as they recorded the fewest alterations and that supported the changes in FA profile analysis. That implies that although CO and PO FMP types had spectral profiles similar to that of freeze-dried raw milk powder (Ejeahalaka & On, 2019a), only the latter appeared to have profiles (i.e., FA and NIRS) nearly as stable as that of WMP under inclement conditions.


absorbance values
0.0
0.5
1.0
1.5
2.0
Wk.0
Wk.1
Wk.2
Wk.3
Wk.4
Wk.5
Wk.6
Wk.7
935nm
1300nm
1463nm
1685nm
1940nm
2010nm
2207nm
2310nm
(a)
1000
1500
2000
2500


absorbance values
0.5
1.0
1.5
935nm
1300nm
1463nm
1940nm
2010nm
2310nm
(b)
1000
1500
2000
2500


absorbance values
0.5
1.0
1.5
2.0
2.5
1000
1500
2000
2500
935nm
1225nm
1300nm
1463nm
1725nm
1940nm
2010nm
2212nm
2310nm
(c)


absorbance values
0.5
1.0
1.5
2.0
2.5
1000
1500
2000
2500
935nm
1225nm
1300nm
1463nm
1530nm
1725nm
1940nm
2010nm
2212nm
2310nm
(d)

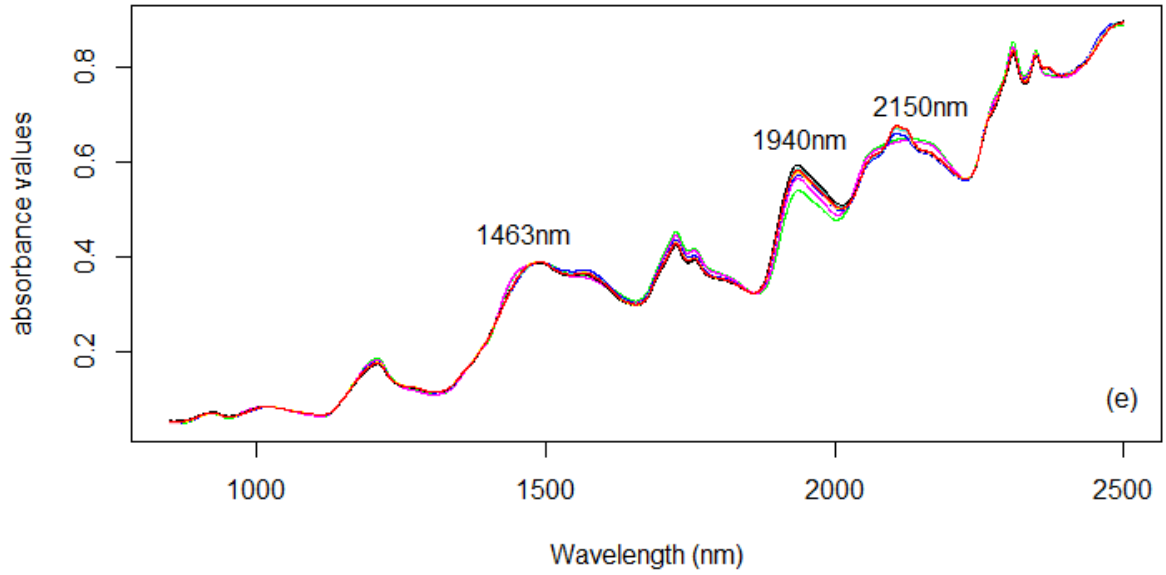


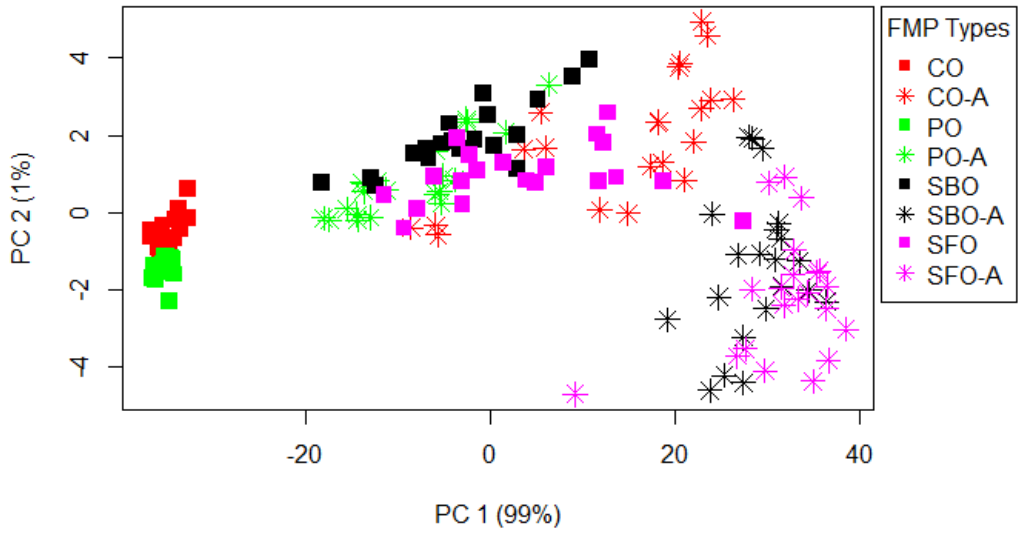
Figure **6.1:** EMSC (extended multiplicative signal correction) pre-processed near infrared spectra of 4 types of fat-filled milk powers (FMP) formulated with 30 % of 4 different types of vegetable oils and stored for 0-7 weeks at 40 °C: (a) coconut oil; (b) palm oil; (c) soya-bean oil; and (d) sunflower oil FMP types; in addition to (e) the spectra of whole milk powder under the same condition. The spectral regions with most of the differential features are indicated on the plots with wavelengths.

6.3.3 PCA fresh and aged fat-filled milk powders characteristics

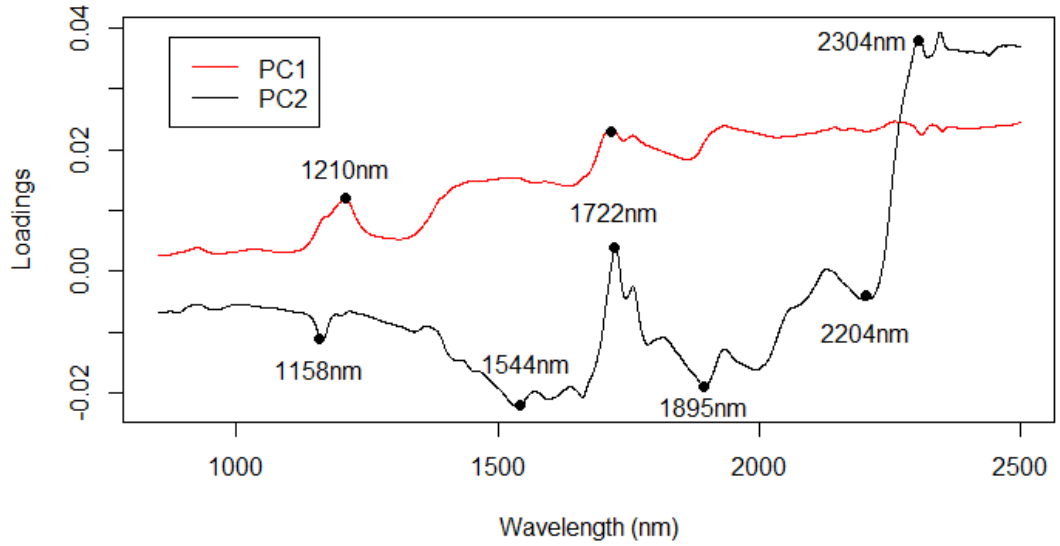
The patterns of variance of the band alterations observed in the spectral profiles of FMP were investigated by plotting the mean-centred PCA of the 164 unique milk samples (i.e., 20 for each of the four fresh FMP types and 21 for each of the four aged FMP types) and then comparing it with that of only the fresh samples that was previously reported (Ejeahalaka & On, 2019a). The score plot of the PCA models for the fresh and aged samples is shown in Figure 6.2a. As can be seen from the figure, the PCA reduced the dimensionality of the data into two principal components, PC 1 and PC 2, which accounted for 99 % and 1 % of the total spectral variance respectively. The score plot showed substantive separation between the fresh FMP types (i.e., CO, PO, SBO and SFO) and their corresponding aged samples (i.e., CO-A, PO-A, SBO-A and SFO-A), thus reflecting the influence of storage on the spectra data. However, there was a significant overlap between the CO-A, PO-A, SBO and SFO FMP types possibly indicating the extent to which the CO and the PO samples may have been altered under the inclement conditions. However, this hypothesis was not supported by the change in FA profile analysis.

Similarly, the SBO-A and SFO-A clustered closely together but appeared well separated from the SBO and SFO FMP types, thus requiring chemometric classification techniques for their characterisation.

The PCA loadings variable plot showing the contributions of the individual wavelengths to PC 1 and PC 2 is depicted in Figure 6.2b. The wavelengths at the loadings plot maxima and minima include: 1158, 1210, 1544, 1722, 1895, 2204 and 2304 nm. These sensitive wavelengths explained the variations observed in the FMP samples during accelerated storage and they were most likely related to C-H band due to lipid modifications (near 1160 nm and 1211 nm for C-H stretch second overtone, 1725 nm for C-H first overtone, and 2300 nm for C-H bending combination), N-H band due to protein modifications (at about 1545 nm and 2205 nm for N-H secondary and primary amines respectively) and O-H band due to water absorption (near 1892 nm for O-H combination band). The comparison of the PCA loadings variable plot for the fresh and aged FMP samples (Figure 6.2b) with that of only the fresh samples (Figure 2 at Ejeahalaka and On (2019a)) showed that PC 2 was significantly altered while that of PC 1 remained fairly stable after the 7 weeks of storage. For instance, the PC 2 of the fresh samples (at t = week 0) had eight strong peaks (i.e., 1216, 1330, 1396, 1589, 1725, 1940, 2096 and 2304 nm) while that of the fresh and aged samples (Figure 6.2b) had only six (i.e., 1158, 1544, 1722, 1895, 2204 and 2304 nm) upon storage, with those at 1330 nm and 1396 nm completely disappearing. Thus, PC 2 correlated more with the aged samples with the loadings weight near 1158 nm, 1544 nm, 1722 nm and 1895 nm having the most pronounced alterations. Grewal et al. (2017) reported in their study of the ultra-high temperature treated whole and skim milk that the samples stored at 40 °C and 50 °C for 28 days showed marked changes in the bands corresponding to conformations of milk lipids and formation of intermolecular β sheet of proteins, indicating protein-lipid interactions and aggregation. On the other hand, the decay of the loadings intensity at about 1158 nm and 1722 nm correlated with those (i.e., 1163 and 1704 nm) found for edible oils upon storage at 60 °C by Wójcicki et al. (2015). However, the wavelength range, 1722 nm to 1895 nm, was chosen as optimal for building the chemometric models as they yielded the most stable predictions.



(a)



(b)

Figure **6.2:** (a) shows the mean centred PCA score plot of the near infrared spectra of the fresh (CO, PO, SBO and SFO) and aged (CO-A, PO-A, SBO-A and SFO-A) fat-filled milk powers (FMP) types. CO, PO, SBO, and SFO represent the coconut oil, palm oil, soya-bean oil and sunflower oil FMP types respectively. Aged samples were stored at 40 °C for t = 7 weeks. (b) illustrates the loadings variable plot of the first two principal components, PC 1 and PC 2, highlighting the sensitive wavelengths.

6.3.4 Multiclass chemometric classification of fresh and aged FMP samples

Table 6.3 shows the performance metrics of the SIMCA models for multiple differentiation of the fresh FMP samples from those that were stored under inclement conditions (40 C and 90 % RH) for a period of 7 weeks. All spectra acquired from t = 0 week (i.e., 20 from each of the 4 fresh FMP types) to t = 7 weeks (i.e., 21 from each of the 4 aged FMP types) were used for the differentiation. Five calibration models were built using four different treatments namely: a) RS with the full spectral range (850 – 2500 nm, 3300 wavelengths); b) KS algorithm with the full spectral range; c) RS with the optimal wavelengths (1722 – 1895 nm, 347 wavelengths) selected from the PCA loadings; and d) KS algorithm with the optimal wavelengths. As can be seen from the table, the fresh samples were firmly differentiated from the aged FMP types (i.e., CO-A, PO-A, SBO-A and SFO-A) without recording any false positives (i.e., specificities = 100 %) irrespective of the calibration modelling treatment that was adopted. That implies that a fresh FMP sample formulated with either coconut, palm, soya-bean or sunflower oil may lose quality, or undergo considerable compositional change over time if subjected to such environmental stress typical of tropical countries. On the other hand, the models developed with the KS algorithm technique performed better than those built with the RS method irrespective of whether variable selection was applied. Overall, the KS algorithm optimised models using the sensitive wavelengths had the best performance metrics in identifying and differentiating the aged FMP samples as they recorded maximum mean RR and E of 100.0 %. The models developed with RS using the full spectra range provided the worst figures of merit (i.e., the mean RR and E values were 61.2 % and 78.2 % respectively). Hence, only the results of the KS algorithm optimised models will be discussed forthwith. As can be seen from the table, the aged-FMP-specific models had sensitivities (bolded diagonal elements) of 100.0 % (i.e., zero false negatives) and that implies that each of the five one-class models was able to recognise its members. All the models had specificity (off-diagonal elements) values of 100.0 % (i.e., zero false positives) and that implies that each of the one-class models was able to correctly reject the test sets that did not belong to the modelled class. Thus, the KS algorithm optimised models recorded mean sensitivity, specificity, RR and E values of 100.0 %, 100.0 %, 100.0 % and 100.0 % respectively across the entire classes. Overall, it appears that the KS algorithm sampling technique may enhance the classification modelling efficiency of FMP where samples were collected on weekly basis over time for the purpose of model development.

Table **6.3:** Performance metrics of the SIMCA models for the classification of fat-filled milk powder (FMP) types[1] formulated at t = 0 week (4 °C) and stored at 40 °C for t = 7 weeks

True class	PC	Test sets[6] assigned into classes					Figures of Merit[7]	
		Fresh	CO-A	PO-A	SBO-A	SFO-A	RR	E
RS with full spectra range[2]								
Fresh	2	**95.0**	100.0	100.0	100.0	100.0	95.0	97.5
CO-A	4	100.0	**80.0**	100.0	100.0	100.0	80.0	89.4
PO-A	5	100.0	100.0	**40.0**	100.0	100.0	40.0	63.2
SBO-A	3	100.0	100.0	100.0	**60.0**	83.3	55.8	75.8
SFO-A	3	100.0	100.0	100.0	40.0	**50.0**	35.0	65.2
KS with full spectra range[3]								
Fresh	3	**95.0**	100.0	100.0	100.0	100.0	95.0	97.5
CO-A	4	100.0	**100.0**	100.0	100.0	100.0	100.0	100.0
PO-A	2	100.0	100.0	**100.0**	100.0	100.0	100.0	100.0
SBO-A	3	100.0	100.0	100.0	**100.0**	100.0	100.0	100.0
SFO-A	4	100.0	100.0	100.0	100.0	**100.0**	100.0	100.0
RS with variable selection[4]								
Fresh	2	**95.0**	100.0	100.0	100.0	100.0	95.0	97.5
CO-A	4	100.0	**100.0**	100.0	100.0	100.0	100.0	100.0
PO-A	2	100.0	100.0	**80.0**	100.0	100.0	80.0	89.4
SBO-A	3	100.0	100.0	100.0	**100.0**	66.7	95.7	91.7
SFO-A	4	100.0	100.0	100.0	80.0	**83.3**	78.3	89.0
KS with variable selection[5]								
Fresh	2	**100.0**	100.0	100.0	100.0	100.0	100.0	100.0
CO-A	2	100.0	**100.0**	100.0	100.0	100.0	100.0	100.0
PO-A	3	100.0	100.0	**100.0**	100.0	100.0	100.0	100.0
SBO-A	3	100.0	100.0	100.0	**100.0**	100.0	100.0	100.0
SFO-A	4	100.0	100.0	100.0	100.0	**100.0**	100.0	100.0

[1]Fresh = control at t = 0 week comprising all FMP types combined (n = 80), prior to trial; CO-A (n =21), PO-A (n =21), SBO-A (n = 21) and SBO-A (n = 21) represent the coconut oil, palm oil, soya-bean oil and sunflower oil FMP types that were aged/stored at 40 °C for t = 7 weeks.
[2]Models developed using the random selection (RS) method and the full spectral range, 850 – 2500 nm (3300 wavelengths) of the calibration sets (n = 124; i.e., 15 for each of the 4 FMP types at t = 0 week and 16 for each of them at t = 7 weeks).
[3]Models developed using Kennard-Stone (KS) algorithm technique and the full spectral range, 850 – 2500 nm of the calibration sets (n = 124).
[4]Models developed using RS and only the sensitive bands, 1722 - 1895 nm (347 wavelengths) selected from the loadings of the principal component analysis.
[5]Models developed using KS algorithm and only the sensitive bands, 1722 - 1895 nm.
[6]Samples (n = 40) of the FMP types not used for building the models made up the test sets. Diagonal elements in bold face represents the sensitivities (%) while the others under the assigned test sets columns indicate the specificity values (%).
[7]RR = reliability rate, E = efficiency and both values are expressed in percent.

6.3.5 PLSR modelling for quantitative prediction of storage time

Storage time, which is one of the key factors that determines the shelf life of milk powders (Romeu-Nadal, Chavez-Servin, Castellote, Rivero, & Lopez-Sabater, 2007), was predicted for the first time for FMP using four different treatments (see Section 6.2.5.2) of the calibration

models and only the aged samples (n = 84 with the calibration set = 64 and test set = 20). The performance statistics is shown in Table 6.4. As can be seen from the table, the iPLS optimised models consistently generated lower errors of cross validation (between 11.8 % and 84.3 %) using a maximum of 5 latent variables and they performed better (i.e., with lower mean RSR and NSE) than the full spectrum models, thus supporting value for variable selection. Hence, only the results for iPLS optimised models will be discussed forthwith. The models developed with RS method using iPLS recorded mean RSR and NSE of 0.19 and 0.95 respectively; whereas those built with KS algorithm technique had values of 0.25 and 0.91 respectively. Thus, the RS method using iPLS had the lowest RSR and the highest NSE values for this study ($R^2p \geq 0.90$, RSR ≤ 0.28, NSE ≥ 0.90) with the CO-A FMP type producing the best results ($R^2p = 0.99$, RSR = 0.08, NSE = 0.99). According to Moriasi et al. (2007), the lower the RSR and the higher the NSE, the better the model prediction performance. This result implied that the length of time for which the FMP containing coconut oil had been in storage under the tropical conditions may be easiest to predict (and that may be an indicator of product freshness and stability) while those containing palm oil may be the hardest. It invariably echoed the relationship between storage time and the product shelf life, and the possibility for prediction. Additional studies of FMP stored under different humidity and temperature conditions would be required for more extensive and accurate modelling.

Table **6.4:** Performance statistics[1] of the PLSR models for estimating the storage time in weeks of fat-filled milk powder types stored under inclement condition (40 °C, 90 % relative humidity) using different sampling methods and variable selection techniques.

Aged FMP[2]	Calibration (n = 64)		Prediction (n = 20)				
	LV	RMSECV	R2p	RMSEP	PBIAS	RSR	NSE
RS with full spectra range[3]							
CO-A	6	0.71	0.94	0.58	-3.63	0.28	0.90
PO-A	8	0.68	0.89	0.75	-8.92	0.36	0.84
SBO-A	4	0.55	0.96	0.55	4.52	0.26	0.91
SFO-A	2	1.46	0.74	1.66	12.96	0.80	0.20
KS with full spectra range[4]							
CO-A	5	0.80	0.98	0.69	6.62	0.46	0.74
PO-A	3	0.90	0.94	0.67	-0.95	0.24	0.93
SBO-A	3	1.02	0.99	0.48	-10.04	0.23	0.93
SFO-A	10	1.51	0.95	0.55	-11.66	0.41	0.79
RS with iPLS[5]							
CO-A	5	0.32	0.99	0.16	0.38	0.08	0.99
PO-A	4	0.60	0.90	0.58	2.14	0.28	0.90
SBO-A	5	0.13	0.99	0.35	-4.62	0,17	0.96
SFO-A	5	0.41	0.95	0.42	0.19	0.20	0.95
KS with iPLS[6]							
CO-A	4	0.44	0.91	0.58	7.17	0.38	0.82
PO-A	5	0.52	0.97	0.73	7.10	0.26	0.91
SBO-A	5	0.16	0.99	0.22	-0.18	0.11	0.99
SFO-A	5	0.54	0.99	0.33	-8.52	0.25	0.92

[1]LV = optimal number of latent variables selected; RMSECV = root mean squared error cross-validation (weeks); R^2_p = coefficient of determination; RMSEP = root mean squared error prediction (weeks); PBIAS = percent bias (%); RSR = RMSE-observation standard deviation ratio calculated as Root Mean Squared Error Prediction/Standard Deviation; NSE = Nash-Sutcliffe efficiency. Spectra were pre-processed with Savitzky-Golay (second derivative with second polynomial order and a window size of 25 points) and the extended multiplicative signal correction method.

[2]CO-A (n = 21), PO-A (n = 21), SBO-A (n = 21) and SBO-A (n = 21) represent the coconut oil, palm oil, soya-bean oil and sunflower oil FMP types that were aged at 40 °C for t = 7 weeks.

[3]Models developed using the random selection (RS) method and the full spectral range, 850 – 2500 nm of the calibration sets (n = 64; i.e., 16 for each of the 4 aged FMP types).

[4]Models developed using Kennard-Stone (KS) algorithm technique and the full spectral range, 850 – 2500 nm of the calibration sets (n = 64).

[5]Models developed using RS and the interval PLS (iPLS) variable selection technique.

[6]Models developed using KS and the iPLS variable selection technique.

6.4 Conclusion

The profile alterations that occurred in aged FMP samples under inclement conditions typical of tropical countries, the export destination for these products, was established through the FA, AA, PCA and the mean spectra analyses. In particular, there was a statistically significant decrease in C18:0 concentration and an apparent increase in the leucine content of the samples for each of the milk fat replacements with vegetable oils upon storage under

environmental stress. Also, the peak NIRS intensity of the O-H functional group increased with storage time; while that of the C-H and C=O functional groups decreased with storage time. Aged CO and SFO FMP types appeared to have unique band alterations near 1685 nm and 1530 nm respectively. Overall, FMP quality alterations may be due to water absorption, lipid and protein modifications; and those containing palm oil may be regarded as the most stable for this study as they recorded the least changes in profile under inclement conditions. Hence, the stability of FMP strongly depends on the type of the incorporated vegetable oils. The multiclass SIMCA models, particularly those built with the combination of KS algorithm technique and the optimal wavelengths, supported the PCA analysis in providing substantive differentiation between the fresh and the aged FMP samples, thus highlighting the harsh consequences of storing the milk powder at high temperature and relative humidity over time. The PLSR models efficiently (NSE ≥ 0.90) predicted the storage time with low errors (RSR ≤ 0.28), indicative of FMP freshness and stability. This study has demonstrated the capability of NIRS to detect compositional alterations in FMP stored under inclement conditions, and thus its utility for rapid and cost-effective screening of FMP for quality assurance purposes.

Chapter 7

Paper V

Monitoring the composition, authenticity and quality dynamics of commercially available Nigerian fat-filled milk powders under inclement conditions using NIRS, chemometrics, packaging and microbiological parameters

Abstract

Fat-filled milk powders (FMP) are inexpensive milk alternatives predominantly exported to developing countries to satisfy growing demands for dairy proteins. Harsh climatic and sanitary conditions, poor border controls and relatively long periods for distribution and storage enhance the inherent vulnerability of FMP to fraud and stability. Rapid, low-cost methods are needed for extensive routine authentication of FMP products. This study investigated, for the first time, the sample integrity and the quality dynamics of 7 Nigerian FMP brands stored for 7 weeks at 40 °C. The prominent melamine and urea absorption peaks were absent, but protein contents were below the permitted limit. The peak absorbance of the O-H functional group increased while the tryptophan contents decreased with storage time. Multiclass analyses differentiated the fresh FMP brands from one another, and from those that were aged. Robust interval-PLS predictions obtained for storage time may be excellent indicators of FMP freshness and stability.

keywords: fat-filled milk powder, near infrared spectroscopy, PCA, multiclass SIMCA screening, interval PLS

7.1 Introduction

Fat-filled milk powders (FMP) are a blend of skimmed milk powder and vegetable fat with a fat and milk protein contents of 28.0 g/100 g and 25.7 g/100 g respectively (Codex Alimentarius Commission, 1999). They are formulated to provide basic nutrition as a cheaper milk alternative for low-income consumers in developing countries. The global demand for FMP is estimated to be growing at 5 % per annum (New Zealand Milk Products, 2019). The European Union (EU), the largest producer of FMP, increased exports by 18.5 % from 2015 to 2018 (CLAL-Dairy Economic Consulting firm, 2019) predominantly to countries in Africa, the Middle East and Russia (ANZ Banking Group, 2018). Indeed, sub-Saharan African countries (including Nigeria, Senegal, Mauritania, Togo and Angola) import 60 % of the EU total FMP exports (European Commission, 2017). The demand for dehydrated milk in Nigeria is very high: 75% of local consumers rely almost entirely on imported dried milk powder (USDA, 2012).

FMP products are shipped to Nigeria from the EU and other temperate countries in 25 kg multiwall bags and then repackaged into 12 g single-serve multilayer laminated sachets for retail distribution within and across the borders of the country (Uppu, 2001). The repackaged FMP, with high fat and protein contents, are exposed to harsh climatic conditions (mean minimum and maximum of the monthly average temperatures and relative humidity from 1951 to 2009 were (16 ºC, 36.6 %) and (35.3 ºC, 85.1%) respectively (Eludoyin, Adelekan, Webster, & Eludoyin, 2014). Sanitary conditions, and a relatively long distribution and storage times (approximately 107 days) (Uppu, 2001) also present challenges in preserving product quality and stability. The repackaging process (performed locally in the facility of the multinational dairy companies in Nigeria) and the high turnover (e.g. Arla Foods reported sales of 11,000 metric-tons in 2012 (Hubschmann, 2013) enhance the inherent vulnerability of FMP to fraud. Nigeria is the international trade hub for the West African sub-region and as such it experiences large number of informal cross-border transactions (Central Bank of Nigeria, 2016) and that complexity may also increase the level of vulnerability of FMP to food fraud threats within the country. In addition, the presence of a limited number of accredited food control laboratories coupled with the haphazard enforcement of food legislation (FAO/WHO, 2003), provide a driver for fraudulent companies to exploit the existing weaknesses to dump substandard FMP products in the country. The potential for tampering, and product

deterioration in undesirable conditions is a strong driver for a rapid, inexpensive method for characterising quality degradation and safety of FMP in circulation in Nigerian markets.

Vibrational spectroscopic instruments acquiring reflectance spectra in the wavelength range of 750 to 2500 nm have been shown to have potential in characterising the quality variances (Ejeahalaka & On, 2019a) and safety (Ejeahalaka & On, 2019b) of FMP products containing different types of vegetable oils as the backbone fats. Near infrared spectroscopy (NIRS) offers high-throughput, cost-effective, non-destructive, user-friendly and fast (one minute or less per sample) analytical measurements. The characteristic near infrared spectra obtained from such measurements can be analysed with chemometric tools in order to understand the spectral properties of the samples and to make modelling decisions without the need for chemical data. To the best of our knowledge, the application of NIRS for routine analysis of commercial samples of FMP has never been reported in the literature.

In this study, we have two goals: (i) to establish the integrity of the single-serve commercial FMP samples in Nigerian markets; (ii) to predict the storage time and the quality alterations that occur in those samples during distribution under the conditions of environmental stress. The baseline variances of the fatty acid (FA) and amino acid (AA) profiles of the fresh and aged commercial samples were also determined to aid understanding of the quality dynamics of the FMP products in circulation in Nigerian markets during prolonged storage and distribution.

7.2 Material and methods

7.2.1 Commercial FMP sample collection

A total of 987 samples (i.e., 141 from each of the 7 FMP brands namely: Three Crowns, Blue boat, Cowbell, Dano, Luna, Olympic and Peak), were purchased from local supermarkets in three of the biggest trading routes and commercial centres (i.e., Lagos, Onitsha and Kano) in Nigeria between 17 July and 18 September 2017. Each of the samples was packed in a 12 g single-serve multilayer laminated branded sachet. Nutritional values were indicated on the sachets, but the type of vegetable fat used in product formulation was not disclosed. The shelf lives of the brands varied and ranged from 12 to 19 months while the production and best before dates were between December 2016 and August 2017, and between March 2018 and November 2018 respectively. In this study, the Three Crowns and Blue Boat FMP brands were denoted as 3-Crown and B-Boat respectively for ease of reference.

7.2.2 Sample integrity assessment

Upon arrival in the laboratory, samples of each of the FMP brands received were opened and immediately screened for chemical and microbial contamination (see Sections 7.2.2.1 and 7.2.2.2 below) at the beginning of the trial.

7.2.2.1 Sample validation with existing calibration models

The near infrared spectra of the fresh (i.e. sachets opened for the first time at t = 0 week, and not subjected to inclement conditions) Nigerian FMP brands (i.e., 3-Crown, B-Boat, Cowbell, Dano, Luna, Olympic and Peak) were measured by first transferring each sachet of the samples into a sealed, sterile, clear glass borosilicate vial and remixing by multiple inversions to ensure homogeneity before subsampling. Measurements were done using a FOSS NIRSystem (model DS 2500F) spectrometer as described previously (Ejeahalaka & On, 2019a). Briefly, 5 g of each of the milk powder samples was placed in a 10-mL DS 2500 ring cup (cup type: 2004) using a sterile powder thief. Then, the sample was scanned at a resolution of 0.5 nm, wavelength accuracy of < 0.05 nm, analysis time of < 1 min using a detector array that was composed of both silicon (850 – 1100 nm) and lead sulfide (1100 – 2500 nm). A total of 140 unique objects (comprising 20 samples from each of the 7 fresh FMP brands) were scanned consecutively in triplicates using the same ring cup without refilling at each scan to minimise sampling error and enhance repeatability. Each spectrum acquired comprised of 3300 absorbance values recorded from 850 to 2499.5 nm averaging 32 scans. The triplicate spectra collected from each sample were averaged and then subjected to chemometric analysis (see Section 7.2.5).

The 20 preprocessed spectra from each of the 7 brands were presented (as test sets) to the true classes (i.e., the calibrated classification models) of the unadulterated (as negative control) and the adulterated (as positive control) FMP types modelled in the previous study (Ejeahalaka & On, 2019b) to predict the extent to which they were different (i.e., their specificities) in terms of the type of vegetable fat (i.e., coconut oil (CO), palm oil (PO), soya-bean oil (SBO) and sunflower oil (SFO)) used in the formulations and the presence of chemical adulterants (i.e., melamine and urea) in the samples respectively. The multiclass analysis performed to assign the test sets to the classes was based on soft independent modelling of class analogy (SIMCA) (see Section 7.2.5.1 for details). The representative spectra of each of the 7 brands were also compared with those of the unadulterated FMP types obtained in the

previous study (Ejeahalaka & On, 2019a) to aid understanding of the type of vegetable fats used in product formulation.

7.2.2.2 Microbial screening for quality evaluation

The background levels of microbes in the FMP brands were assessed using two process hygiene indicators (i.e., the aerobic plate count (APC) and the sulphite-reducing clostridia spores count (SRCs); and two food safety parameters (i.e., *Salmonella* and *Cronobacter* (formerly *Enterobacter) sakazakii*). All analytical procedures were performed independently by Eurofins Laboratory Christchurch using standard methods. In brief, the APC was evaluated using the milk plate count agar according to the International Standard Organisation (ISO) 4833-2 protocol (ISO, 2013). The SRCs count was evaluated based on ISO 15213 using the differential reinforced clostridia agar (ISO, 2003). *Salmonella* screening was achieved through the process of pre-enrichment, selective enrichment, selective plating and serological confirmation using the ISO 6579 protocol (ISO, 2002). The presence of *C. sakazakii* in the samples was evaluated through the process of pre-enrichment, selective enrichment, isolation, confirmation and biochemical characterisation using the ISO/TS 22964 : IDF/RM 210 protocol (ISO/IDF, 2006).

7.2.3 Storability assessment of Nigerian FMP under tropical conditions

Twenty-one FMP samples from each of 7 brands (total n=147), which were packed individually in multilayer single-serve laminated branded sachets, were used for the assessment. The oxygen transmission rate (OTR) of the sachets was measured at 23 °C and 0 % relative humidity (RH) following the ASTM D3985 standard. The water vapour transmission rate (WVTR) of the sachets was measured at 38 °C and 90 % RH according to ASTM F1249 protocol. Before the commencement of the trial, each of the laminated sachets containing the samples was carefully inspected to ensure that there was no leakage. Then, the sealed FMP sachets were stored in a Contherm incubator (Contherm Scientific Ltd, New Zealand), away from direct sunlight, at a temperature of 40 ± 1 °C and RH of 90 ± 5 % for a period of 7 weeks. The FMP samples were placed on a dry plastic tray, positioned above a temperature-controlled reverse-osmosis purified water-bath, within a large sealable plastic container with lid, in order to create the saturated environment. The temperature and RH data were recorded in real time using sensors and data loggers. High target values of temperature (40 °C) and RH (90 %) were chosen for the trial to replicate the tropical conditions obtainable in Nigeria and to accelerate

potential deteriorative changes in FMP within the shortest possible time. The stored/aged FMP samples from the 7 brands were denoted as 3-Crown-A, B-Boat-A, Cowbell-A, Dano-A, Luna-A, Olympic-A and Peak-A. Sampling involving NIRS measurements (see Section 2.2.1) was done weekly in triplicates and each of the sachets was measured once. Chemical analyses (i.e., FA and AA profiling) were also performed in triplicates at the beginning of the trial (t = 0 week) on fresh samples and at the end of the 7 weeks of storage.

7.2.4 Chemical analyses of the Nigerian FMP samples

7.2.4.1 Fatty acid determination

The methods employed for the determination of FA concentrations have been described in detail previously (Ejeahalaka & On, 2019a). In brief, a one-step methylation process was done by adding 900 µL of heptane and 4.0 mL of 0.5M NaOH/dried methanol to Kimax tubes containing 0.15 g of the FMP sample and 100 µL of internal standard (C21:0 ester). Then, the tubes were carefully vortexed and incubated in water-bath at 50 °C for 15 min before the addition of 2.0 mL of heptane and 2.0 mL of distilled water. The mixture was centrifuged at 1500 g for 5 min to separate the heptane layer of the FA esters. Small amounts of anhydrous sodium sulphate were subsequently added to remove residual water and a subsample of the recovered extract called FA methyl esters (FAME) was stored in a vial at − 20 °C until GC analysis.

The GC analysis was carried out by injecting 1.0 µL of the extracted FAME onto a Varian CP7420, tailor-made fused silica capillary column with a length of 100 m, internal diameter of 0.25 mm, and film thickness of 0.20 µm, using the AOC-20i auto-sampler fitted to a Shimadzu GC-2010 gas chromatograph, as described previously (Ejeahalaka & On, 2019a). The FA concentrations obtained from the assay were expressed as g of fatty acid per 100 g of total fatty acids. The changes in the FA concentrations and the lipid indices of atherogenicity (IA) and thrombogenicity (IT) (Ulbricht & Southgate, 1991) of the fresh and aged samples were evaluated to aid understanding of the FA dynamics of Nigerian FMP under environmental stress.

7.2.4.2 Amino acid profiling

The profiling of the AA composition of the samples was done in two steps: acid hydrolysis and high-performance liquid chromatograph (HPLC) analysis, as described in detail previously (Ejeahalaka & On, 2020). In brief, the acid hydrolysis was performed by adding 5.0 mL 6 N of

hydrochloric acid into each of the three tubes containing 0.1 g of the FMP sample and 10.0 µL 0.5 M of internal standard (amino-butyric acid). Then, the tubes were carefully vortexed and purged with nitrogen (oxygen free) for 30 seconds before heating at 110 °C for 20 hours in a block heater (Ratek, Australia). The tubes were centrifuged (GeneVac Scientific, USA) at a maximum temperature of 47 °C and the resulting dried residue was reconstituted into a 50ml volumetric flask. The hydrolysate solution was then filtered through a 0.45 µm syringe filter into a 2 mL vial and kept frozen at – 20 ºC in triplicates until HPLC analysis.

The HPLC analysis was carried out by injecting 11.0 µL of pre-column derivatised sample onto an ACE C-18 column (ACE, UK) of dimension 150 x 4.6 mm and particle size 3 µm using an auto-sampler with thermostat fitted to an Agilent 1100 series HPLC system with EZChrom Elite software (Agilent Technologies, Walbronn, Germany) as described previously (Ejeahalaka & On, 2020). The derivatised amino acids were detected by fluorescence detector and the separation was completed in 36 min. The standard curve was established by using individual amino acids standards. The results obtained were expressed in g of amino acids per 100 g of FMP samples.

7.2.5 Chemometric modelling of Nigerian FMP brands

Chemometric multivariate analyses were performed on the raw spectra of the 287 unique objects (comprising 20 samples from each of the 7 fresh FMP brands; and 21 from each of the 7 aged FMP brands) using the R software version 3.5.1 (R Core Team, 2018). The mean spectra of the FMP brands were first plotted for each of the 7 weeks of storage after preprocessing with the extended multiplicative signal correction (EMSC) technique (Stark & Martens, 1996) to highlight the subtle changes in profiles (Ejeahalaka & On, 2020) that occurred as a result of exposure to inclement conditions. Then, the principal component analysis (PCA) was performed on the mean-centered raw spectra as the first classification step to reduce dimensionality and to identify any natural groupings among the different FMP brands. The PCA loadings was subsequently plotted to visualize the contributions of the original variables and to identity the dominant wavelengths that contain most of the spectral information. To differentiate between the fresh and the aged Nigerian FMP brands in order to understand the real impact of environmental stress on the samples and to predict the storage time, the raw spectra were preprocessed using a Savitzky-Golay (SG) second derivative algorithm (window size = 55 points, second-order polynomial fit), followed by an EMSC to preserve higher moments and to remove baseline effects. The preprocessed spectra were divided into a

training set (for building the models) and a testing set (for independent valuation of the models) in the ratio of 4:1 (Ejeahalaka & On, 2019a) using the random selection (RS) method and the Kennard-Stone (KS) algorithm as described previously (Ejeahalaka & On, 2020).

Statistical analyses were performed by paired *t*-test as previously described (Ejeahalaka & On, 2020) to determine whether there was a change in the FA and AA contents of the seven Nigerian FMP brands that were received at t = 0 week (4 °C) and stored for t = 7 weeks at 40 °C (significance level $p < 0.05$). PCA analyses were also conducted to help understand the pattern of changes that occurred in the FA and AA contents of the FMP brands as a result of storage under inclement conditions (i.e., 40 °C and 90 % RH).

7.2.5.1 FMP differentiation by supervised pattern recognition

SIMCA, a supervised pattern recognition technique, was used as the second classification step to differentiate among the fresh FMP brands and to assign the aged samples to the class models. To do this, seven one-class SIMCA models were built (one for each of the fresh brands) using the SG + EMSC preprocessed calibration sets (i.e., n = 105 comprising 15 samples from each of the 7 fresh FMP brands) to test the ability of each of them to recognise the test samples from the modelled classes (i.e., n = 35 comprising 5 samples from each of the 7 fresh FMP brands); and those belonging to the aged FMP brands (i.e., n = 147 since all the 21 aged samples from each of the 7 brands were presented to the models for validation). The sampling was done using the RS method and the KS algorithm technique. The calibration models were built using only the sensitive wavelength range selected from the PCA loadings to simplify model interpretation and to optimise the predictions. The goodness of the classification rules was determined through internal validation using the leave-one out (LOO) cross validation method (Ejeahalaka & On, 2019a). Prediction performances were assessed using only the specificity values when the aged FMP brands were assigned to the seven class models whereas the sensitivity and specificity values were used as the figures of merit for the assignation of the fresh FMP test sets to the models. Sensitivity was calculated as the fraction of the samples belonging to the modelled class which was correctly accepted or recognised by the model; whereas specificity was defined as that fraction of samples not belonging to the modelled class that was correctly rejected by the model (Oliveri & Downey, 2012).

7.2.5.2 FMP storage time predictive algorithm

Partial least square regression (PLSR) technique was used to construct the predictive models for the storage time (in weeks) of the aged Nigerian FMP brands based on the measured SG +

EMSC preprocessed spectral data. Two different sampling treatments (i.e., the RS method and the KS algorithm) were applied. A total of 112 samples (i.e., 16 from each of the 7 aged brands) were used as training sets (with 35 as test sets (i.e., 5 from each of the 7 aged brands)) to build 7 PLSR models for the aged FMP brands. Each of the models was built with only the optimal spectral intervals selected using the interval partial least square (iPLS) modelling procedure proposed by Nørgaard et al. (2000). The number of latent variables (LV) that corresponded to the lowest root mean squared error of cross validation (RMSECV) was chosen as optimal and it was determined through internal validation using the LOO cross validation method. Then, the root mean squared error of prediction (RMSEP) was estimated through external validation of the models using independent test sets that were not used for the calibrations. The overall goodness of fit of the models was determined using the RMSE-observation standard deviation ratio (RSR) (i.e., the ratio of RMSEP to standard deviation of the reference/measured data or the reciprocal of the residual prediction deviation) (Moriasi et al., 2007), and the Nash-Sutcliffe efficiency (NSE) (i.e., the relative magnitude of the residual variance compared to the measured data variance) (Nash & Sutcliffe, 1970). These metrics were used for similar assessment in our previous study (Ejeahalaka & On, 2020). Predictions were only considered to be good if RSR < 0.5 and NSE was between 0.0 and 1.0 (Moriasi et al., 2007).

7.3 Results and discussion

7.3.1 FMP authenticity and microbial quality

Table 7.1 shows the barrier properties, the microbial quality characterisation and the SIMCA validation of Nigerian brands with existing calibration models of unadulterated (Ejeahalaka & On, 2019a) and adulterated (Ejeahalaka & On, 2019b) FMP types. As can be seen from the table, the OTR and the WVTR values for the 7 branded commercial flexible sachets ranged from 0.08 to 19.51 cc m^{-2} day^{-1} and from 0.08 to 0.64 g m^{-2} day^{-1} respectively. The laminated multilayer sachets for Dano FMP brand had the highest OTR and WVTR while those for Luna and Peak samples recorded the lowest values respectively. Thus, the Luna FMP sachets provided the best barrier for oxygen while those for Peak brand had the best barrier for water vapour. Generally, low OTR (i.e., high oxygen barrier) and low WVTR (i.e., high water vapour barrier) are the prerequisite conditions for preserving the product quality throughout its whole lifecycle (Schmid et al., 2012).

The APC for the 7 commercial FMP ranged from 20 to 800 cfu/g, well within the 5000 cfu/g limit based on Codex Alimentarius principles specified for such products in this region (World

Food Programme, 2014) and indicative of the process hygiene quality and the level of adherence to good manufacturing practices for these brands. Cowbell and Dano FMP had the lowest bacterial populations (i.e., 20 cfu/g) whereas 3-Crown brand had the highest count with 800 cfu/g. In addition, all FMP brands had SRCS count of < 10 cfu/g except for 3-Crown which recorded a value of 10 cfu/g. However, *E. sakazakii* and *Salmonella* spp. were not detected in any of the brands studied, indicating sound production values.

Profile comparisons shown in Appendix C (Fig. C1) revealed that the raw representative spectra of the Nigerian commercial brands, particularly 3-Crown, B-Boat, Cowbell, Dano and Peak, looked similar in every respect and overlapped substantially with those of the unadulterated CO and PO FMP types described in the previous study (Ejeahalaka & On, 2019a). The NIRS profiles of the Nigerian FMP samples were readily distinguished from FMP samples produced in our laboratory using SBO and SFO (Ejeahalaka & On, 2019a), but more closely resembled spectra of CO and PO FMP types. Fatty acid analysis (Section 7.3.2.1 below) indicate that PO formed a major component of the Nigerian products but cannot preclude the use of alternative oils such as copra that may have been used to co-formulate product. There are other variables in FMP production (Hansen, 1980) that may also have influenced the NIRS profiles of Nigerian FMP. Coupled with the non-commercial production process we used to formulate our own FMP samples (Ejeahalaka & On, 2019a), it is a testament to the validation performance (Table 7.1) of our SIMCA calibration models that the Nigerian samples were not assigned to any of the unadulterated or adulterated (i.e., with melamine or urea) FMP training sets (i.e., specificities = 100 %) based on our previous work (Ejeahalaka & On, 2019a, 2019b). Besides, the prominent absorption peaks for melamine (at about 1955 and 2000 nm) and urea (at about 1315 and 2000 nm) (Ejeahalaka & On, 2019b) were not sighted in any of the spectra from the 7 brands in the comparisons shown in Fig. C4. Although specific tests for melamine and urea were not undertaken on the Nigerian FMP samples due to cost limitations, our NIRS results (i.e., profile comparisons in Fig. C4 and SIMCA validation in Table 7.1) indicate them to be unadulterated, a finding concordant with the low level of microbial contaminants determined (see above) indicating good manufacturing practice.

Table **7.1:** Sample integrity assessment and the barrier properties of 7 commercial brands of fat-filled milk powders (FMP) obtained from the Nigerian markets

Barrier properties of Nigerian branded commercial flexible sachets							
	Branded single serve commercial flexible sachets						
Properties[1]	3-Crown	BlueBoat	Cowbell	Dano	Luna	Olympic	Peak
OTR (cc m^{-2} day^{-1})	5.98	0.17	1.15	19.51	0.08	3.35	0.89
WVTR (g m^{-2} day^{-1})	0.49	0.10	0.58	0.64	0.14	0.54	0.08
Microbial quality characterisation of Nigerian FMP samples							
	Nigerian single serve FMP brands						
Parameter[2]	3-Crown	BlueBoat	Cowbell	Dano	Luna	Olympic	Peak
APC (cfu/g)	800	40	20	20	400	150	40
E. sakazakii	Absent	Absent	Absent	Absent	Absent	Absent	Absent
Salmonella	Absent	Absent	Absent	Absent	Absent	Absent	Absent
SRCS count	10	<10	<10	<10	<10	<10	<10
SIMCA validation of Nigerian samples with known unadulterated and adulterated FMP types							
	Specificity values of Nigerian FMP test sets[4] on assignation into classes						
True class[3]	3-Crown	BlueBoat	Cowbell	Dano	Luna	Olympic	Peak
CO	100.0	100.0	100.0	100.0	100.0	100.0	100.0
PO	100.0	100.0	100.0	100.0	100.0	100.0	100.0
SBO	100.0	100.0	100.0	100.0	100.0	100.0	100.0
SFO	100.0	100.0	100.0	100.0	100.0	100.0	100.0
COM	100.0	100.0	100.0	100.0	100.0	100.0	100.0
COU	100.0	100.0	100.0	100.0	100.0	100.0	100.0
POM	100.0	100.0	100.0	100.0	100.0	100.0	100.0
POU	100.0	100.0	100.0	100.0	100.0	100.0	100.0
SBOM	100.0	100.0	100.0	100.0	100.0	100.0	100.0
SBOU	100.0	100.0	100.0	100.0	100.0	100.0	100.0
SFOM	100.0	100.0	100.0	100.0	100.0	100.0	100.0
SFOU	100.0	100.0	100.0	100.0	100.0	100.0	100.0

[1]Oxygen transmission rate (OTR) was measured at 23°C/0% relative humidity (RH) according to ASTM D3985 whereas water vapour transmission rate (WVTR) was measured at 38°C/90% RH (ASTM F1249).

[2]APC = aerobic plate count; and SRCS count = sulphite reducing clostridial spores count (cfu/g).

[3]One-class models for inhouse unadulterated (i.e., CO (coconut oil), PO (palm oil), SBO (soya-bean oil) and SFO (sunflower oil)) and adulterated (i.e., COM (coconut oil with melamine), COU (coconut oil with urea), POM (palm oil with melamine), POU (palm oil with urea), SBOM (soya-bean oil with melamine), SBOU (soya-bean oil with urea), SFOM (sunflower oil with melamine), and SFOU (sunflower oil with urea)) FMP types were built as previously reported (Ejeahalaka & On, 2019a).

[4]A total of 140 samples (i.e., 20 samples from each of the 7 brands) were assigned to each of the true class models as test sets. Specificity values describe the percentage of samples not belonging to each of the true classes (cf. Ejeahalaka and On, 2019a).

7.3.2 Quality dynamics of Nigerian FMP brands during storage

7.3.2.1 Fatty acid profiling and alterations

The FA concentrations of the 7 commercial FMP brands, including the alterations that occurred in their values during storage under inclement conditions (i.e., 40 °C, 90 % RH), are shown in Table 7.2. As can be seen from the table, the short-chain (i.e., C4:0 and C6:0) and the medium-chain (i.e., C8:0 and C10:0) FA were omitted because they were either absent or present in trace quantities, indicating that the milk fats were not of animal origin. Bovine milk fat contains about 10.9 % of C4:0 to C10.0 FA with butyric acid (C4:0) contributing approximately 4.4 % by weight (Lindmark Månsson, 2008). The gross concentrations of FA in the FMP brands were approximately 52.5 % saturated and 45.8 % unsaturated, reflecting the use of PO as the milk fat replacement in their formulations. This assertion is supported by the fact that PO has a balanced FA composition in which the level of saturated fatty acids (SFA) is almost equal to that of unsaturated fatty acids (UFA) (Siew, 2011). Moreover, similar weights were obtained for SFA and UFA in PO FMP type in the previous study (Ejeahalaka & On, 2019a). The SFA in the FMP brands ranged from 51.8 % to 54.6 %, with C16:0 as the most dominant, contributing about 88.5 % of those FA. Olympic brand contained the highest amounts (54.6 %) of SFA and Cowbell the least (51.8 %). The UFA in the FMP brands ranged from 42.4 % to 46.8 %, with C18:1c9 as the most dominant, contributing about 78.9 % of those FA. Peak brand contained the highest amounts (46.8 %) of UFA and Olympic the least (42.4 %). The absence of the *Capra* FA (i.e., C6:0, C8:0 and C10:0) and the dominant presence of C16:0 (for SFA) and C18:1c9 (for UFA) coupled with the balanced weights of SFA and UFA in the 7 FMP brands suggested that CO was at least not the main vegetable fat used for the products formulation. This is because the incorporation of CO as the milk fat replacement has been shown in the previous study (Ejeahalaka & On, 2019a) to result in FMP containing predominantly medium chain FA (especially C12:0) with SFA of about 91.7 % in addition to low quantities of C18:1c9 and C18:2c9,12. The monounsaturated FA in the 7 commercial FMP ranged from about 33.9 % to 37.5 %, with Peak brand having the highest weight (37.5 %) and Olympic, the least. The linoleic (LA, C18:2c9,12) and the alpha-linolenic (ALA, C18:3 c9,12,15) acids contents of the FMP brands ranged from 8.4 % to 10.1 % and 0.1 % to 0.3 % respectively, resulting in LA to ALA ratios of about 27.8:1 to 63:1. These ratios are very high compared to those reported for cow milk (1.1:1), goat milk (2.0:1) and cow infant formula (10.0:0) by Prosser, Svetashev, Vyssotski, and Lowry (2010). A very high LA (i.e., omega-6) to ALA (i.e., omega-3) ratio of about 15.0:1 to 16.7:1 (optimal values varied from 1.0:1 to 4.0:1) has been shown (Simopoulos,

2002) to promote the pathogenesis of many diseases including cardiovascular disease, cancer, and inflammatory and autoimmune diseases. The IA and IT for the FMP brands ranged from 1.1 to 1.4 and 2.1 to 2.3 respectively. These values were nearly the same as those (i.e., IA = 1.2, IT = 2.4) previously reported (Ejeahalaka & On, 2019a) for PO FMP type and the brands were generally less atherogenic than whole milk powder (IA = 1.7) (Vargas-Bello-Pérez et al., 2017). Hence, although the spectral comparisons above (Fig. C4) showed that the profiles of the 7 commercial brands resembled those of CO and PO FMP types, their FA concentrations were more like that of the latter. The PCA score plot of the Nigerian brands in Fig. C5(a) showed that the FA profiles of 3-Crown, B-Boat and Dano were similar (since they were grouped together in the same quadrant), and they were inversely correlated with those of Olympic FMP which was located in a diagonally opposed quadrant. In the same vein, the FA concentrations of Cowbell, Luna and Peak FMP brands were grouped together in the lower right quadrant (Fig. C5(a)) and as such were correlated (i.e., had some similarities in FA profiles). The biplot of the FA profiles (Fig. C5(b)) shows that C16:0, C20:0, C18:1 c9 and C18:2 c9,12 were associated with 3-Crown and Peak; C18:3 c9,12,15 with Cowbell and Luna; and C12:0, C14:0, C18:0 and C16:1 c9 with Olympic FMP brand.

The FMP brands responded differently to environmental stress after the 7 weeks of accelerated storage. However, the SFA concentrations of the samples generally decreased while their UFA contents increased upon storage except for B-Boat FMP which followed an opposite trend. Cowbell and Peak FMP recorded the least, albeit minimal, alterations in SFA and UFA concentrations while Luna and Olympic brands had the most changes. Consequently, the IA and IT of Cowbell and Peak brands respectively remained nearly unchanged after the 7 weeks of storage. Overall, the changes recorded in the FA profiles of Nigerian FMP brands under inclement conditions were not statistically significant according to the paired *t*-test results shown in Table C1 (Appendix C).

Table **7.2:** Changes in fatty acid concentrations of 7 commercial brands of fat-filled milk powder samples obtained from the Nigerian markets and stored for 7 weeks at 40 °C and 90 % relative humidity

Fatty Acids[1]	3-Crown	BlueBoat	Cowbell	Dano	Luna	Olympic	Peak
FA (g/100g FA) before storage at t=0 week							
C12:0	0.16	0.25	0.27	0.19	0.20	3.69	0.21
C14:0	0.84	0.91	1.09	0.96	0.94	2.63	0.91
C16:0	47.02	47.24	46.31	47.63	46.70	43.92	46.48
C18:0	3.89	3.70	3.81	3.66	3.80	4.10	3.98
C20:0	0.30	0.27	0.28	0.28	0.30	0.27	0.29
∑SFA	52.21	52.37	51.76	52.72	51.94	54.61	51.87
C16:1 c9	0.12	0.14	0.15	0.14	0.13	0.18	0.13
C18:1 c9	36.13	36.76	36.87	35.49	36.79	33.69	37.37
C18:2 c9,12	10.00	8.87	9.24	10.08	9.46	8.35	9.06
C18:3 c9,12,15	0.22	0.14	0.29	0.20	0.34	0.22	0.19
∑UFA	46.47	45.91	46.55	45.91	46.72	42.44	46.75
LA/ALA	45.45	63.36	31.86	50.40	27.82	37.95	47.68
IA	1.09	1.11	1.09	1.13	1.08	1.37	1.08
IT	2.17	2.22	2.13	2.22	2.11	2.32	2.15
Changes (%) in FA after 7 weeks of storage							
C12:0	1.58	1.26	-2.57	-0.19	13.05	0.40	1.86
C14:0	0.76	0.70	-1.74	-0.85	1.99	-2.74	0.89
C16:0	-0.13	0.08	-0.12	-0.16	-0.59	0.12	-0.07
C18:0	-0.60	0.38	1.30	-0.08	0.95	-2.45	-0.55
C20:0	-1.12	0.07	1.60	-1.17	-1.60	-1.62	-1.03
∑SFA	-0.15	0.12	-0.06	-0.17	-0.39	-0.20	-0.09
C16:1 c9	1.40	-0.41	1.03	-0.28	2.13	-3.74	1.13
C18:1 c9	-0.10	0.12	-0.01	0.11	1.38	0.37	0.05
C18:2 c9,12	1.01	-1.04	0.21	0.33	-3.12	0.92	-0.05
C18:3c9,12,15	1.64	-1.65	-1.34	0.38	-5.10	-2.35	0.52
∑UFA	0.15	-0.11	0.03	0.16	0.42	0.45	0.04
IA	-0.22	0.24	-0.30	-0.36	-0.77	-0.83	-0.03
IT	-0.34	0.25	-0.03	-0.32	-0.65	-0.60	-0.14

[1]Each fatty acid (FA) concentration represents the mean of triplicate measurements; SFA = saturated fatty acids; UFA = unsaturated fatty acids; LA = C18:2 c9,12; ALA = C18:3 c9,12,15; IA = index of atherogenicity, IT = index of thrombogenicity and both indices were estimated according to Ulbricht & Southgate (1991).

7.3.2.2 Amino acid profiling and alterations

Table 7.2 shows the AA concentrations of the commercial FMP brands and the corresponding changes that occurred in their values after 7 weeks of storage at 40 °C and 90 % RH. As can be seen from the table, the concentrations of the non-essential AA (i.e., alanine, aspargine, glutamate, serine and tyrosine) were not presented. The essential amino acids (EAA), including the conditionally EAA (i.e., arginine, proline, cysteine, glycine and taurine), ranged from 9.3 to 14.0 g/100g of milk, representing 58.8 % of the total AA with the Luna brand having the highest value and Olympic the least. Proline, leucine and lysine were the most abundant EAA and they

accounted for about 10.1 %, 8.7 % and 8.6 % respectively of the total AA in the 7 commercial brands studied. Cowbell brand had the highest concentration of proline whereas Luna and 3-Crown FMP had the highest amounts of leucine and lysine respectively. Branched-chain AA (i.e., leucine, isoleucine and valine) constituted about 18.7 % of the total AA while the sulfur AA (i.e., cysteine and methionine) made up approximately 3.0 % of them. Among the brands, Olympic and Peak had the lowest cysteine concentration while B-Boat had the highest. The total AA concentrations of the 7 FMP ranged from 15.6 to 23.7 g/100g of milk, with Olympic brand uniquely having the lowest value. Consequently, as the AA compositional analysis can be used to assess the protein levels (Kambhampati, Li, Evans, & Allen, 2019), it can be concluded that the total AA recorded for each of the 7 brands was lower than the recommended (Codex Alimentarius Commission, 1999) protein content (i.e., 25.7 g/100 g of milk) for FMP. The PCA score plot of the Nigerian brands in Fig. C6(a) showed that the AA profiles of Dano and Peak before storage (at t = 0 week) were similar and they were inversely correlated with those of B-Boat and Olympic. Similarly, the AA profiles of 3-Crown, Cowbell and Luna FMP brands at t = 0 week were correlated since they were grouped together in the same quadrant.

Milk proteins undergo changes during the storage of dry powders (Higgs & Boland, 2014) and the single serve multilayer laminated commercial FMP samples appeared to be no exception. The PCA score plot of the AA concentrations (Fig. C6) provided substantive differentiation between the fresh and the aged Nigerian brands, in stark contrast to the degree of separation obtained in the plot of their FA profiles (Fig. C5). For instance, some of the correlations that existed between the fresh FMP brands on the AA PCA score plot (as discussed above) were lost and new relationships were established post storage. Hence, it appeared that the inclement storage conditions (40 °C, 90 % RH) had a more pronounced influence on the AA than on the FA profiles of the Nigerian FMP commercial brands, an observation analogous to that seen in our previous study (Ejeahalaka & On, 2020). As can be seen from Table 2, almost all the FMP brands had an increase in histidine concentration (though not statistically significant (Table C2)) and a decrease in tryptophan, lysine, leucine, EAA and the total AA contents after the 7 weeks of accelerated storage. Statistically significant (Table C2, Appendix C) changes in concentrations, averaging 17.8 % and 4.5 %, were observed in tryptophan ($t(6) = -5.07$, $p = 0.002$) (a hydrophobic AA) and lysine ($t(6) = -3.48$, $p = 0.013$) contents respectively across the 7 FMP brands studied. In addition, the Nigerian brands recorded a statistically significant decrease in total AA ($t(6) = -2.56$, $p = 0.043$) contents. The loss of lysine contents

after the 7 weeks of storage may have resulted from the interactions of proteins and lactose under high temperature and RH (i.e., Maillard reactions) or through the formation of isopeptide bond (Higgs & Boland, 2014). On the other hand, it appeared that the large losses of tryptophan contents signalled the occurrence of protein damage in the samples as a result of lipid-protein co-oxidation (i.e., lipid oxidation and subsequent co-oxidation with protein) since the FMP brands are a complex milk model of high lipid and high protein (Wazir et al., 2019). Previous research (Estévez & Xiong, 2019) has shown that the consumption of food products that have undergone lipid oxidation and protein co-oxidation (i.e., oxidised protein) can lead to serious health risks such as oxidative stress and related diseases. Overall, Cowbell and Olympic brands recorded the least alterations in the contents of EAA and the total AA respectively.

Table **7.3:** Changes in amino acid concentrations of 7 commercial brands of fat-filled milk powder samples obtained from the Nigerian markets and stored for 7 weeks at 40 °C and 90 % relative humidity

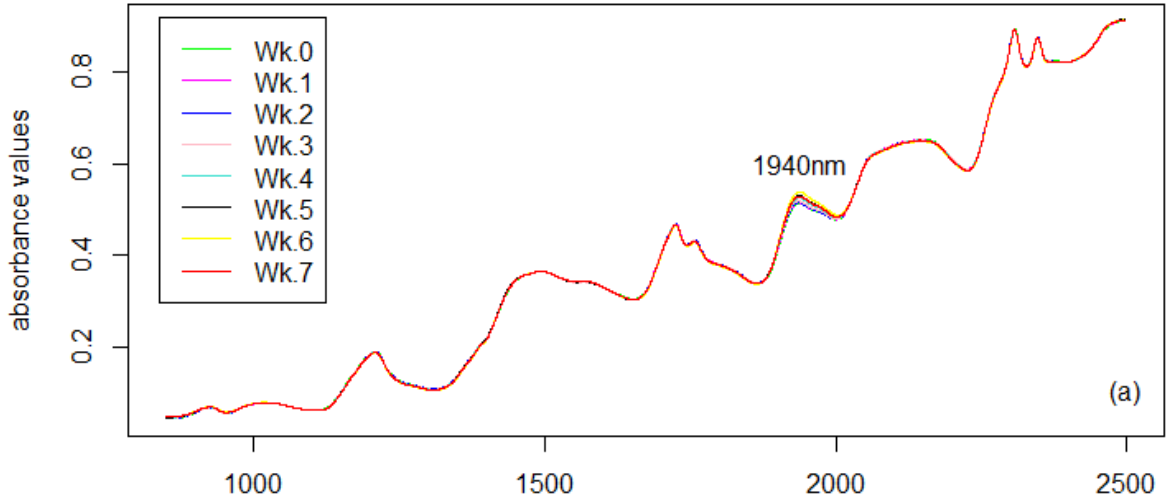
Amino Acids[1]	3-Crown	BlueBoat	Cowbell	Dano	Luna	Olympic	Peak
AA (g/100g of milk) before storage at t=0 week							
Lysine	2.03	1.53	2.01	1.94	2.00	1.32	1.91
Threonine	0.96	0.86	0.97	0.92	0.96	0.73	0.91
Tryptophan	0.20	0.14	0.18	0.17	0.22	0.20	0.16
Leucine	2.01	1.67	1.98	1.86	2.11	1.41	1.84
Isoleucine	1.05	0.91	1.07	1.00	1.10	0.76	0.96
Valine	1.21	1.01	1.23	1.17	1.28	0.84	1.13
Methionine	0.52	0.41	0.52	0.46	0.54	0.31	0.49
Phenylalanine	0.97	0.76	0.97	0.91	1.02	0.64	0.90
Histidine	0.87	0.72	0.91	0.92	0.85	0.54	0.87
Arginine	0.79	0.62	0.82	0.81	0.79	0.55	0.76
Proline	2.25	1.90	2.34	2.23	2.52	1.59	2.09
Cysteine	0.15	0.18	0.17	0.16	0.17	0.14	0.14
Glycine	0.39	0.32	0.40	0.36	0.40	0.29	0.36
Taurine	0.01	0.01	0.01	0.01	0.01	0.01	0.01
∑EAA	13.41	11.04	13.57	12.91	13.98	9.32	12.52
∑AA	23.05	18.69	23.06	21.96	23.72	15.58	21.50
Changes (%) in AA after 7 weeks of storage							
Lysine	-3.46	0.63	-7.30	-9.61	-4.52	-3.58	-3.39
Threonine	1.08	-2.97	-1.74	-3.55	-1.47	1.75	1.31
Tryptophan	-18.17	2.12	-20.06	-30.45	-19.34	-19.78	-18.85
Leucine	-2.84	-5.36	-1.47	-0.67	-6.56	0.03	-0.28
Isoleucine	0.20	-5.15	-3.12	-0.02	-5.41	3.45	1.26
Valine	1.54	-3.54	-1.82	-1.51	-4.90	3.27	1.93
Methionine	-1.74	-0.89	0.01	7.35	-6.51	11.91	-1.50
Phenylalanine	-2.85	3.64	4.41	-1.39	-4.76	0.41	0.85
Histidine	13.36	22.63	5.80	-11.63	3.06	8.81	12.22
Arginine	3.64	13.61	1.84	-15.54	-4.20	-4.37	5.06
Proline	-2.23	-6.78	6.90	-0.38	-8.75	-5.55	5.55
Cysteine	9.99	-24.49	-13.18	9.71	-5.91	7.03	2.70
Glycine	-2.21	1.11	3.81	0.49	-0.31	4.58	0.85
Taurine	-2.57	3.71	3.95	-14.98	4.75	-14.47	1.02
∑EAA	-0.51	-0.74	-0.15	-3.91	-5.23	-0.27	1.72
∑AA	-0.84	-2.72	-1.76	-5.10	-4.18	-0.71	0.77

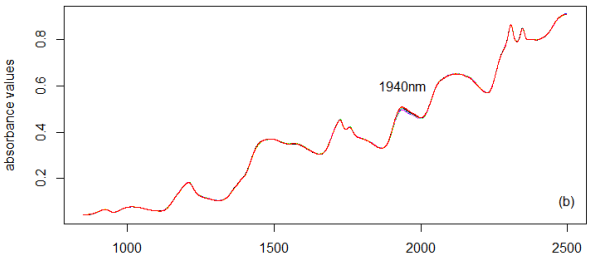
[1]Each amino acid (AA) concentration represents the mean of triplicate measurements. Only the details of the essential amino acids (EAA) were presented.

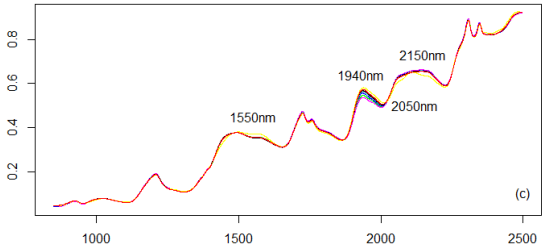
7.3.2.3 Near infrared spectra alterations

The deviations recorded in the fingerprints of the FMP brands across the 7 weeks of accelerated storage were investigated using the EMSC pre-processed near infrared spectra shown in Fig. 7.1. As can be seen in the figure, spectra alterations, which occurred where the absorbance values changed with the storage times, were only visible at 1940 nm for 3-Crown,

B-Boat, Dano, Luna, Olympic and Peak brands; and at the wavelengths 1550, 1940, 2050 and 2150 nm for Cowbell FMP. According to Workman and Weyer (2008), the wavelengths 1550, 1940, 2050 and 2150 nm may be related to the first overtone of N-H stretching, O-H stretching and HOH bending combination, N-H stretching combinations, and C-H stretch and C=O stretch respectively. Thus, the near infrared spectra of each of the brands exhibited the unique water combination band at 1940 nm reflecting the possibility of water migration into the samples on exposure to environmental stress. Overall, the spectra of the fresh and aged samples were superimposed at most of the wavelengths and as such there were minimal visible alterations after the 7 weeks of accelerated storage, thus requiring the use of chemometric tools for their possible differentiation.


absorbance values
0.2
0.4
0.6
0.8
Wk.0
Wk.1
Wk.2
Wk.3
Wk.4
Wk.5
Wk.6
Wk.7
1940nm
(a)
1000
1500
2000
2500


absorbance values
0.2
0.4
0.6
0.8
1940nm
(b)
1000
1500
2000
2500


0.2
0.4
0.6
0.8
1550nm
1940nm
2150nm
2050nm
(c)
1000
1500
2000
2500

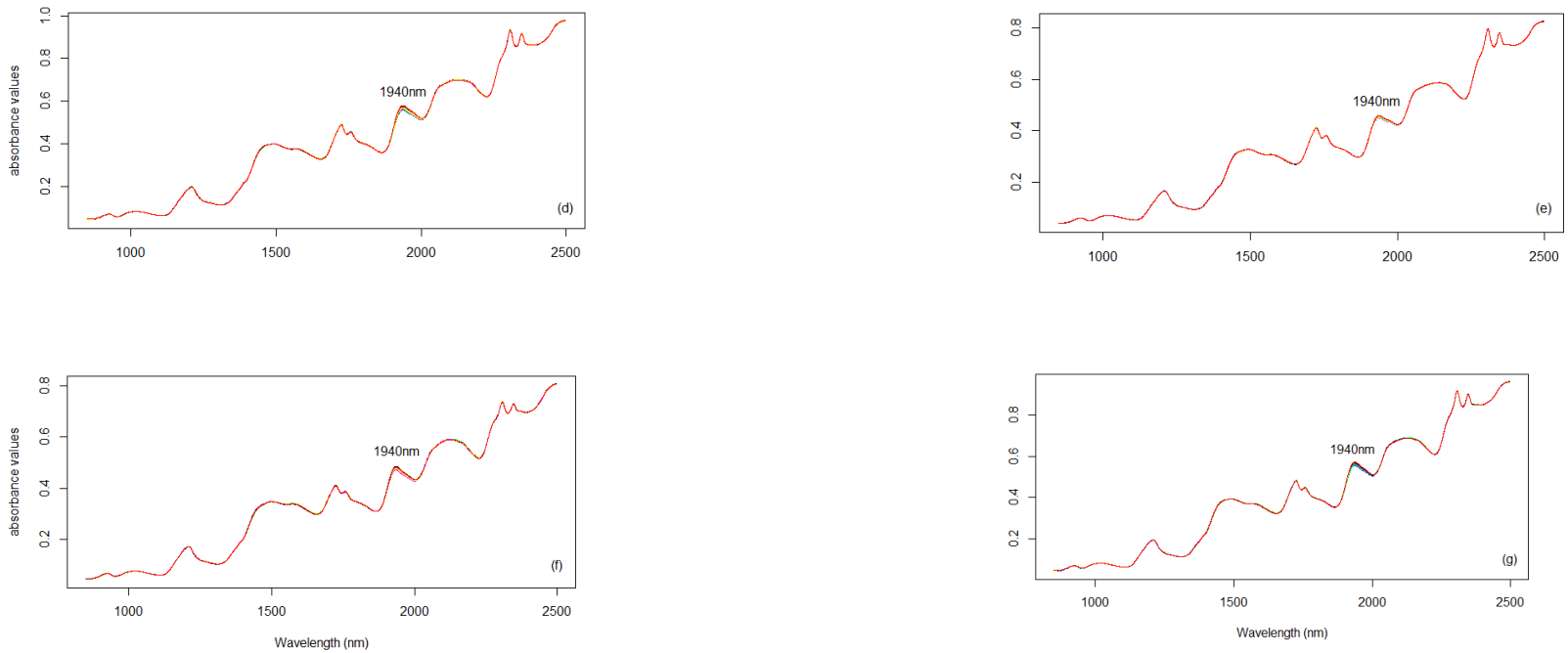
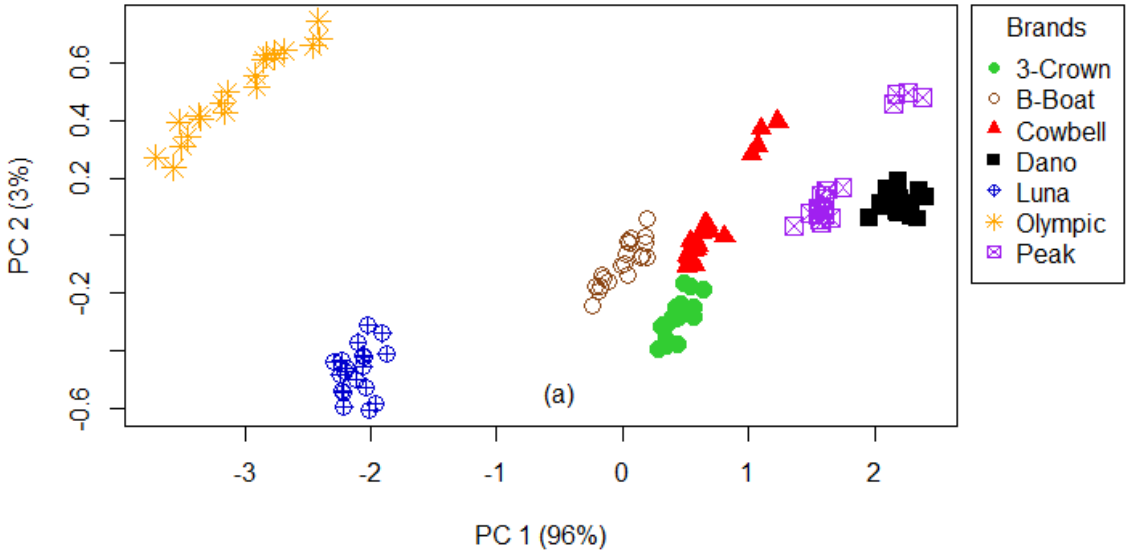


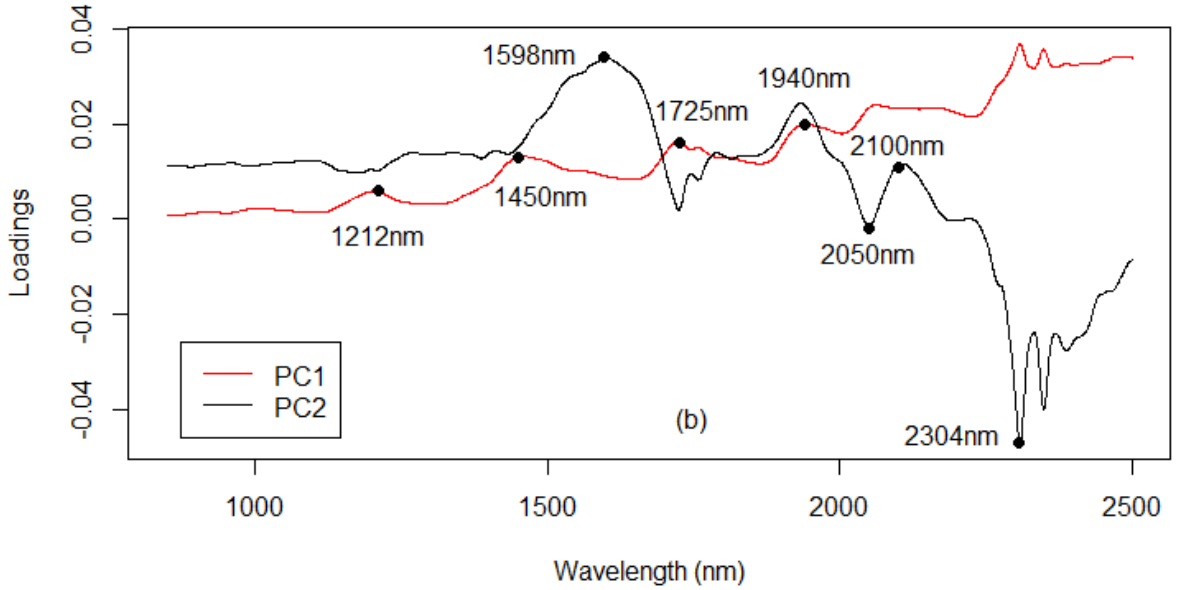
Figure **7.1:** EMSC (extended multiplicative signal correction) pre-processed near infrared spectra of 7 commercial brands of fat-filled milk powder (FMP) samples obtained from Nigerian markets and stored for 0-7 weeks at 40 °C and 90 % relative humidity: (a) 3-crown; (b) blue boat; (c) cowbell; (d) dano; (e) luna; (f) olympic; and (g) peak FMP types.

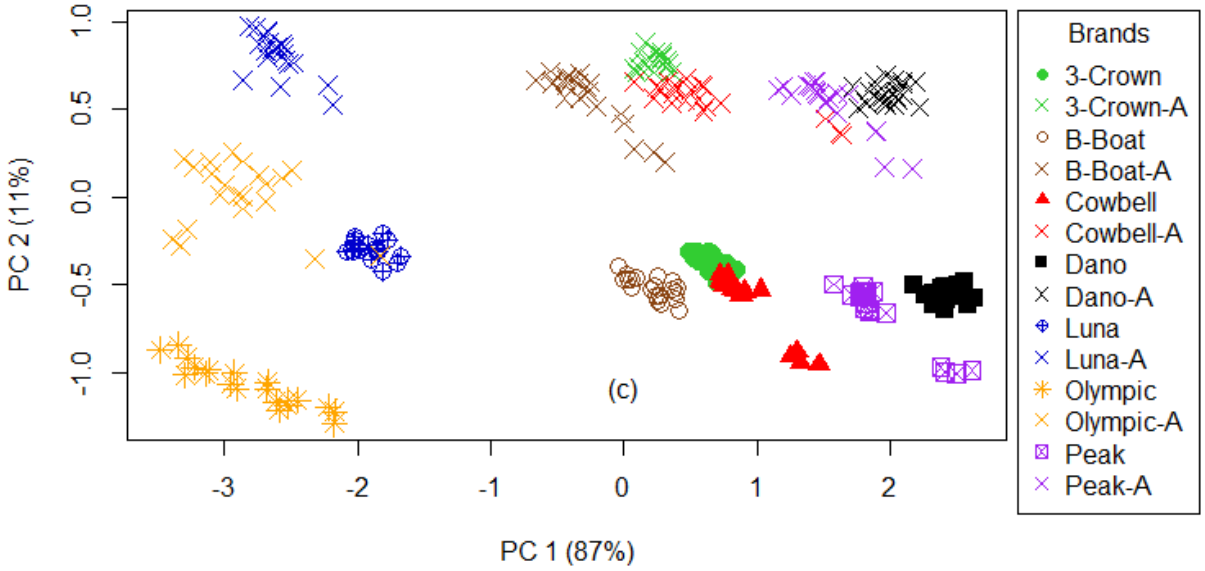
7.3.2.4 PCA spectra alterations pattern description

The similarities and the differences, including the pattern of alterations captured in the spectra (850 - 2500 nm, 3300 wavelengths) of the FMP brands (n = 287) under environmental stress (i.e., 40 °C and 90 % RH) were investigated using PCA. Consequently, the 3300 spectral variables were reduced into two significant principal components, PC 1 and PC 2, which retained 98 % of the cumulative variance in the data, with PC 1 corresponding to the direction of maximum variance as shown in the scores/loadings plots in Fig. 7.2. For the fresh samples, the scores plot (Fig. 7.2a) appeared to show 7 substantive clusters (one for each brand), with PC 1 effectively separating Luna and Olympic from the rest of the FMP brands. That probably showed that Luna and Olympic FMP had profiles that differed from those of the other brands, and these differences were reflected in the spectra comparisons analyses (Fig. C4) previously discussed in Section 7.3.1. However, Luna and Olympic FMP were negatively correlated since they were positioned on the opposite sides of the scores plot. This finding was partially supported by the divergence in AA profiles (Table 7.3). On the other hand, 3-Crown, Cowbell, B-Boat, Dano and Peak brands probably shared similar characteristics since they were located close to each other in the plot. However, Dano and Peak FMP were positively correlated since they were positioned next to each other. In the same vein, B-Boat and Cowbell brands were positively correlated, although the former was located close to the origin of the scores plot and as such can be said to have average characteristics. The loadings variable plot (Fig. 7.2b) for the fresh samples showed that 8 wavelengths (i.e., 1212, 1450, 1598, 1725, 1940, 2050, 2100 and 2304 nm) made the largest contributions in the clustering of the FMP brands in the scores. The bands at 1212 and 1725 nm may be due to the second and the first overtones of C-H stretching of methylene respectively (Workman & Weyer, 2008); whereas that at 2304 nm may be related to the combinations between elongation and deformation of methylene groups (Kaddour et al., 2006). The absorption peaks at 1450 and 1940 nm were due to the -OH groups of water whereas the bands around 1598 and 2050 nm correlated with the -NH groups in protein (Workman & Weyer, 2008). Also, the 2100 nm band was associated with the O-H bending and C-O stretching combination of polysaccharides (Workman & Weyer, 2008).
On the other hand, the post-storage (at t = 7 weeks) scores plot (Fig. 7.2c) offered substantive separation between the fresh and the aged FMP brands, with the PC 1 and PC 2 accounting for 87 % and 11 % of the total spectral variance respectively. Thus, PC 2 accounted for less of

the cumulative variance but was more effective in separating the aged FMP samples from the fresh brands. The post-storage loadings variable plot (Fig. 7.2d) in comparison to that for the fresh FMP brands (Fig. 7.2b) showed a wide separation between PC 1 and PC 2 with the profile pattern of the latter totally reversed, thus acknowledging the possible occurrence of compositional alterations in the sample matrices as a result of storage under inclement conditions. To provide more details, the pre-storage (at t = 0 week) loadings variable plot (Fig. 2b) had 8 maxima/minima at 1212, 1450, 1598, 1725, 1940, 2050, 2100 and 2304 nm whereas the post-storage loadings (Fig. 7.2d) had only 6 of those strong absorption intensities at 1212, 1450, 1725, 1940, 2050 and 2304 nm, with the bands near 1598 and 2100 nm completely disappearing. The post-storage extinction of the bands at 1598 and 2100 nm, which correlated with the -NH groups in protein and the O-H bending and C-O stretching combination of polysaccharides respectively (Workman & Weyer, 2008), signified the possible occurrence of Maillard reaction (i.e., reaction of lactose and milk proteins) defects in the FMP brands as a result of the inclement storage conditions. In addition, the loadings at 2304 nm, which was previously associated with the C-H bond absorption in fats (Ejeahalaka & On, 2019a), recorded a sharp decay in band intensity signifying the progression of lipid oxidation process. According to Zamora and Hidalgo (2011), lipid oxidation and Maillard reaction are dependent pathways and as such the carbonyl compounds produced in the former will compete with carbohydrate-derived carbonyls for amino compounds to produce carbonyl-amine reaction products with either beneficial or harmful properties. On the other hand, the loadings weights at 1450 and 1940 nm, which correlated with the O-H functional group, grew in intensity with storage time reflecting the increased moisture content of the FMP brands and the concomitant occurrence of lactose crystallisation with the associated stickiness and caking phenomena (Huppertz & Gazi, 2016). That probably explains the visible presence of large lumps in some of the FMP samples after the 7 weeks of storage. These results supported the assertion by Chopovda et al. (2016) that humidity enhances moisture permeation into the sachets thereby lowering the glass transition temperature, and pushing the powder closer to lactose crystallization. Thus, most of the differential features between the loadings variable plots before and after storage of the FMP brands may have resulted from the deteriorative effects of lipid-protein co-oxidation and lactose crystallisation. However, to further distinguish between the 7 commercial FMP brands, the wavelength range 1725 – 2050 nm (i.e., 651 wavelengths) was chosen as optimal for building the chemometric models as it yielded better predictions with higher interpretability.


Brands
3-Crown
B-Boat
Cowbell
Dano
Luna
Olympic
Peak
PC 2 (3%)
0.6
0.4
0.2
-0.2
-0.6
-3
-2
-1
0
1
2
PC 1 (96%)
(a)


Loadings
0.04
0.02
0.00
-0.02
-0.04
1598nm
1940nm
1725nm
2100nm
1450nm
1212nm
2050nm
2304nm
PC1
PC2
(b)
1000
1500
2000
2500
Wavelength (nm)


PC 2 (11%)
1.0
0.5
0.0
-0.5
-1.0
-3
-2
-1
0
1
2
PC 1 (87%)
(c)
Brands
3-Crown
3-Crown-A
B-Boat
B-Boat-A
Cowbell
Cowbell-A
Dano
Dano-A
Luna
Luna-A
Olympic
Olympic-A
Peak
Peak-A

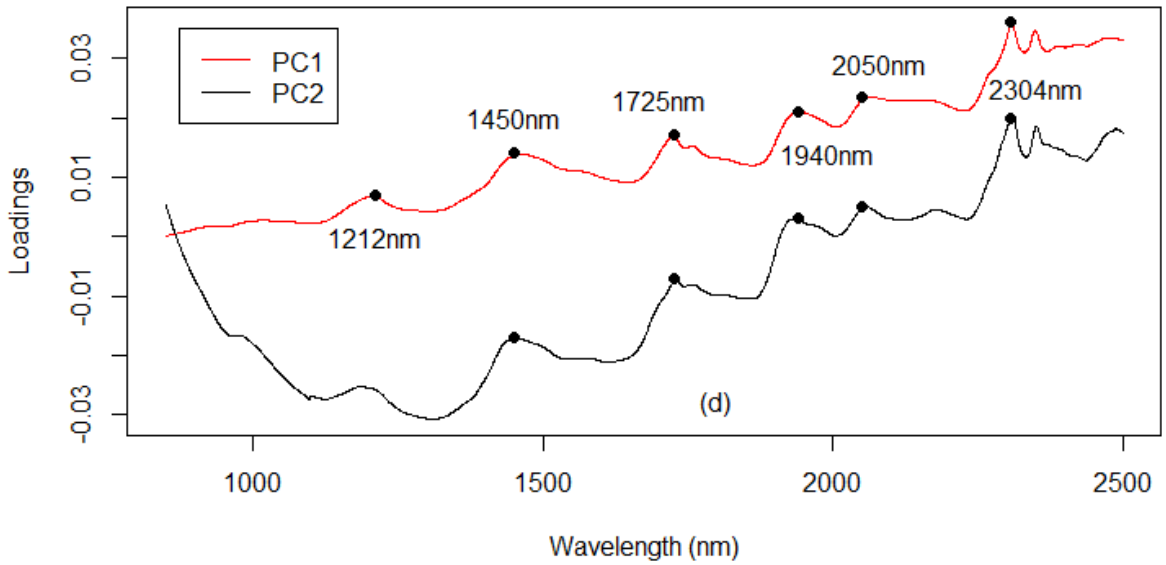


Figure **7.2:** shows: (a) the mean centred PCA score plot of the near infrared spectra of 7 commercial brands (i.e., 3-crown, b-boat, cowbell, dano, luna, olympic and peak) of fresh (at t=0 week) fat-filled milk powders (FMP) obtained from Nigerian markets; (b) the loadings variable plot of the first two principal components, PC 1 and PC 2, of fresh FMP samples highlighting the sensitive wavelengths; (c) the mean-centred PCA score plot of the fresh and aged (at t=7 weeks) (i.e., 3-crown-A, b-boat-A, cowbell-A, dano-A, luna-A, olympic-A and peak-A) Nigerian FMP samples; and (d) the loadings variable plot of PC1 and PC2 of the fresh and aged FMP samples showing the sensitive wavelengths.

7.3.2.5 Multiclass SIMCA analysis for FMP brands differentiation

Seven one-class SIMCA calibration models were built (one for each of the fresh FMP brands at t = 0 week) using only the sensitive wavelengths (1725 – 2050 nm, 651 wavelengths) of the EMSC + SG preprocessed spectra of 105 training sets (i.e., 15 for each of the 7 fresh FMP brands) selected by either the RS method or the KS algorithm technique. Then, the test sets of the fresh brands (n = 35; i.e., 5 from each of the 7 fresh brands) and all the aged FMP (n = 147; i.e., 21 from each of the 7 aged brands) were independently presented to each of the 7 calibration models to determine their abilities to recognise them as their members (i.e., sensitivity) or reject them as not belonging to their classes (i.e., specificity). The sensitivity and specificity values obtained from the multiclass analyses using the RS method and the KS algorithm technique are shown in Table 7.4 and Table C3 respectively. Overall, the KS algorithm optimised model (Table C3) performed slightly better in classifying the FMP brands and as such will only be discussed forthwith. As can be seen from the top of the table under SIMCA classification of fresh Nigerian FMP brands, the 7 one-class models (i.e., the true classes) were able to classify the 35 objects (i.e., the 5 samples from each of the 7 fresh brands) with no false positives, yielding specificity values of 100 %. That implies that the 7 fresh commercial FMP brands were distinct, and successfully differentiated using NIRS. Similarly, most of the calibration models had sensitivities of 100 % (except for Olympic which had value of 80 %) and as such will be able to recognise their own members. These results probably indicated that the fresh Nigerian FMP brands were of high homogeneity.

On the other hand, the SIMCA classification of aged Nigerian FMP brands (at the middle of Table 7.4) showed that the 7 one-class models were able to classify the 147 objects (i.e., all the 21 samples from each of the 7 aged brands) with no false positives, yielding specificity values of 100 %. That implies that the true classes were completely different from the aged FMP brands. These results supported the PCA analyses and they suggest that although the commercial FMP brands were repackaged into single-serve multilayer laminated sachets, the fresh samples differed from those in storage or retail distribution that were subjected to diverse environmental stresses typical of the tropical countries such as Nigeria.

Table **7.4:** Performance statistics of the multiclass SIMCA and PLSR models for Nigerian fat-filled milk powders (FMP) received at t = 0 week and stored for 7 weeks at 40 °C and 90 % relative humidity

SIMCA classification[1] of fresh (t = 0 week) Nigerian FMP brands							
	Sensitivity and specificity of fresh test sets (n = 35) on assignation into classes						
True class[2]	3-Crown	BlueBoat	Cowbell	Dano	Luna	Olympic	Peak
3-Crown	80.0	100.0	100.0	100.0	100.0	100.0	100.0
BlueBoat	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Cowbell	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Dano	100.0	100.0	100.0	80.0	100.0	100.0	100.0
Luna	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Olympic	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Peak	100.0	100.0	100.0	100.0	100.0	100.0	100.0
SIMCA classification of aged (t = 7 weeks) Nigerian FMP brands							
	Specificity values of all aged samples (n = 147) on assignation into classes						
True class	3-Crown	BlueBoat	Cowbell	Dano	Luna	Olympic	Peak
3-Crown	100.0	100.0	100.0	100.0	100.0	100.0	100.0
BlueBoat	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Cowbell	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Dano	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Luna	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Olympic	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Peak	100.0	100.0	100.0	100.0	100.0	100.0	100.0
PLSR predictions[3] of the storage time in weeks of Nigerian FMP brands							
	FMP brands aged/stored at 40 °C for 7 weeks						
Metrics[4]	3-Crown	BlueBoat	Cowbell	Dano	Luna	Olympic	Peak
R^2p	0.93	0.98	0.99	1.00	0.92	0.83	0.94
RMSECV	0.61	0.36	0.10	0.27	0.20	0.65	0.32
RMSEP	0.63	0.35	0.27	0.07	0.95	0.78	0.50
RPD	3.27	5.99	7.64	28.28	2.19	2.67	4.19
RSR	0.31	0.17	0.13	0.04	0.46	0.37	0.24
NSE	0.88	0.97	0.98	1.00	0.74	0.83	0.93

[1]Classification models were built using only the sensitive bands, 1725 – 2050 nm, of the preprocessed calibration sets (n = 105; i.e., 15 from each of the 7 fresh brands) selected based on random sampling.

[2]True class represents the one-class model built for each of the brands using the calibration sets of the fresh samples at t = 0 week.

[3]Prediction models were built using only the sensitive intervals (i.e., interval PLS) of the preprocessed calibration sets (n = 112; i.e., 16 from each of the 7 aged brands) selected based on random sampling.

[4]R^2p = coefficient of determination for prediction (n = 35); RMSECV = root mean square error of cross validation (weeks); RMSEP = root mean square error of prediction (weeks); RSR = RMSE-observation standard deviation ratio calculated as 1/RPD (ratio of prediction to deviation) (optimal RSR < 0.5); NSE = Nash-Sutcliffe efficiency.

7.3.3 PLSR modelling for storage time prediction

Seven PLSR models were developed using only the iPLS optimal intervals of the EMSC + SG preprocessed spectra for the prediction of storage time for the 7 commercial FMP brands. Two different sampling treatments (i.e., the RS method and the KS algorithm) of the calibration models were investigated and only the aged Nigerian samples (n = 147; with the calibration set = 112 (i.e., 16 from each of the 7 aged brands) and the test set = 35 (i.e., 5 from each of the 7 aged brands)) were used. The figures of merit obtained are shown in Table 7.4 and Table C3. The optimised models developed with the RS method (Table 7.4) recorded mean RSR and NSE of 0.25 and 0.90 respectively; while those built with the KS algorithm technique (Table C3) had values of 0.36 and 0.79 respectively. Thus, the RS method had the lowest RSR and the highest NSE values, and these results are consistent with those obtained in our previous study (Ejeahalaka & On, 2020). Consequently, for this study, the iPLS models based on the RS method efficiently (NSE ≥ 0.74) predicted the storage time with low errors (RSR≤ 0.46), with the Dano FMP having the best predictions (RSR = 0.04, NSE = 1.00) and the Luna brand the worst (RSR = 0.46 (optimal < 0.50), NSE = 0.74). These results inversely correlated with the barrier properties (Section 7.3.1) of Dano and Luna FMP brands. The root mean squared error (RMSE) recorded was < 0.95 % and the maximum difference found between RMSEC and RMSEP was 0.75 %, indicating that the models were robust enough to make good predictions. Luna FMP brand recorded the maximum RMSEP of 0.95 % which was less than the 1.04 % value calculated as half the standard deviation of the reference storage time that was desired for good quantitative calibration models (Ejeahalaka & On, 2019b). Thus, the storage time for Luna FMP will be hardest to predict whereas that for Dano brand will be easiest. These results showed that the length of time for which the commercial FMP brands had been exposed to inclement storage condition can be estimated with good efficiency, and the predictions may be excellent indicators of product freshness and stability.

7.4 Conclusion

The near infrared spectra of the 7 Nigerian commercial brands did not contain the prominent melamine and urea absorption peaks and they shared similar characteristics with the fingerprints of known unadulterated CO and PO FMP types. However, SIMCA validation was successful and the FA analysis showed that the commercial FMP brands contained predominantly palm oil as the milk fat replacements. Thus, the FMP brands were of acceptable

chemical and microbial integrity, although their protein contents appeared to be below the permitted limit. The dynamics of the qualities of the FMP brands during storage or retail distribution at high temperature and RH were reflected in the FA, AA, PCA and SIMCA analyses. In particular, the tryptophan contents decreased significantly and the peak NIRS intensities at 1598 and 2100 nm became extinct after storage, signifying the occurrence of lipid-protein co-oxidation and Maillard reaction defects respectively. Also, the peak absorbance of the O-H functional group increased with storage time, thus lowering the glass transition temperature and leading to lactose crystallization. Overall, most of the differential features observed in the quality parameters of the FMP brands during storage under inclement conditions resulted from the deteriorative effects of Maillard reaction, lipid-protein co-oxidation and lactose crystallisation. The SIMCA multiclass models supported the PCA analysis in differentiating the fresh FMP brands from one another, and from those that were stored under tropical conditions. Robust PLSR models were developed for storage time, and the predictions obtained ($R^2p \geq 0.83$, $RPD \geq 2.19$, $RSR \leq 0.46$, $NSE \geq 0.74$) may be excellent indicators of FMP freshness and stability. This study has demonstrated the potentials of NIRS to screen for chemical adulterants and to monitor the quality dynamics of commercial FMP brands in storage or retail distribution under diverse environmental stresses typical of the tropical countries such as Nigeria.

Chapter 8

Conclusions and perspectives

8.1 Summary of investigations

Global dairy producers are moving steadily towards producing differentiated products aimed at niche markets with greater emphasis on maximising resource utilisation and meeting consumer preferences. Consequently, FMP are produced to increase the sales of underutilised SMP, while cows with specialised characteristics and requirements are aggregated into different herds for targeted nutritional management to reduce the milk FA that are often associated with negative health effects, offering improved economic returns. Rapid methods are therefore desirable for independent verification of product quality and origin to support validation and traceability of such products. NIRS appears to be ideal for objective characterisation of FMP products for quality assurance and for the segregation of raw milk from individual cows for the manufacture of high value niche products, yet we have not found evidence of its application in these regards. To effectively bridge these gaps in knowledge, we formulated hypotheses to determine: (a) the extent to which NIR spectral analysis can be related to FA profiles of individual raw milk samples, and discriminate milk from different cow herds; (b) whether NIR spectral analysis can be used to differentiate FMP products with differing backbone fats used in their manufacture; (c) whether NIR analysis can reliably discriminate FMP products originating from different countries and brands; (d) whether NIR spectral analysis can be used to profile the quality of FMP samples stored under conditions of environmental stress; (e) the efficacy (specificity, sensitivity) of NIR spectral profiling for detection of adulterants, and the conclusions that can be drawn from the examination of FMP products from Nigeria; and (f) the relationship, if any, between microbial content of samples and potential adulterants. Investigations commenced with data mining and the use of NIRS to segregate freeze-dried raw milk samples from individual dairy cows from mixed breeds, which were aggregated into different herds under the same or differing feeding regimes (see Chapter 3, Paper I). The NIRS profiles of the raw milk samples were analysed with chemometric tools and correlated with the reference crude protein and the FA phenotypes of individual cows to assess the efficacy of the predictions and the utility of the approach in differentiating milk quality between herds. Then, three proportions (i.e., 10 %, 20 % and 30 %) of four types of vegetable oils (i.e., CO, PO, SBO and SFO) were blended with SMP as the backbone fats to

formulate FMP samples under laboratory conditions and their impacts and baseline variances were investigated using the FA and NIRS profile analyses coupled with chemometrics (see Chapter 4, Paper II). Thereafter, the freshly formulated FMP samples were deliberately spiked with differing amounts (0.01 – 16.00 %) of melamine and urea to examine the efficacy of NIRS to detect and quantify such chemical adulterants with no nutritional value that can potentially be used to adulterate the milk powder; and that will be valuable in the instances of fraud (see Chapter 5, Paper III). To do this, three multilevel chemometric models were built using HMS as a tool to overcome the potential complexity of the sample matrices resulting from the mixture of the backbone fats (i.e., the vegetable lipids) and the adulterants in the spiked FMP samples. Then, the need to acquire superior knowledge of the storage behaviour of FMP under inclement conditions prompted a stability trial to be conducted for 7 weeks at 40 °C and 90 % RH (see Chapter 6, Paper IV). Multiclass analyses were performed on the fresh and aged FMP samples using the full spectrum and the sensitive NIR wavelengths to characterise the quality alterations that occurred due to environmental stress in the midst of the interactions between the milk proteins and the backbone vegetable fats. The storage time was predicted indicative of FMP freshness and the changes in FA and AA profiles were evaluated to understand the influence of the type of the incorporated vegetable oil on milk powder stability during prolonged storage. With a solid understanding of the typical profiles of unadulterated and adulterated model FMP, commercial samples covering 7 FMP brands were collected from 3 main trading routes in Nigeria to test the applicability of all our qualitative and quantitative models that have been so far developed under laboratory conditions as mentioned above (see Chapters 4 to 6, Papers II to IV). Hence, the sample integrity (i.e., chemical adulterants and microbial screening) and the quality dynamics of 987 single-serve Nigerian FMP samples were investigated for 7 weeks at 40 °C and 90 % RH using NIRS coupled with new and existing chemometric models (see Chapter 7, Paper V). The baseline variances of the FA and the AA profiles of the fresh and aged commercial samples were also analysed to aid understanding of the quality dynamics of the FMP products in circulation in Nigerian markets during prolonged storage and distribution.

8.2 Summary of findings and conclusions

The qualitative and quantitative analyses conducted in this research have demonstrated that NIRS can reliably segregate freeze-dried raw milk from individual cows/herds for product innovation, providing quality assurance and traceability benefits, but may not be adequate in correlating their FA profiles in accordance with the Beer's law. Consequently, poor predictions of the FA concentrations of raw milk from individual cows were recorded due to matrix effects, although our analyses were conducted on freeze-dried samples which may not be feasible under on-farm conditions for the use of NIRS as a quick method for *in situ* measurements. Nevertheless, the study provided useful, novel methodology for discriminating raw milk at herd level for the benefit of manufacturers of on-farm made-to-order high value products.

We found that although bovine milk lipids can be replaced with cheaper indigenous vegetable oils (i.e., PO, SBO and SFO) to produce milk alternatives that may benefit arterial health, the use of CO for fat replacement resulted in FMP that were more atherogenic and thrombogenic. Chemical and spectral profile similarities were recorded between FMP of similar fat saturation (e.g., CO and PO) making it difficult for them to be adequately identified and differentiated only by visual inspection and FA profile analysis. However, NIRS coupled with chemometrics, particularly the PLS-DA models developed with the NIR spectral profiles, provided substantive differentiation (i.e., zero false negative and positive values with mean efficiency of 100 %) among the FMP products containing different types and proportions of vegetable fats in their formulation irrespective of the spectra pre-treatment used. PO FMP had spectral profiles and lipid indices closest to raw milk but SBO and SFO FMP types, with significantly higher NIRS absorption intensities, were the most desirable options in relation to coronary heart disease incidence.

The spiking of the fresh samples of the FMP types (i.e., CO, PO, SBO and SFO) with different concentrations (i.e., 0.01 to 16.00 %) of melamine and urea appeared to have resulted in some compositional alterations in the sample matrices. Consequently, the ordering pattern of the NIRS absorption intensities were altered with the adulterated PO FMP type recording the least absorbance values. The spectral differences between the adulterated and unadulterated FMP types appeared generally indiscernible to the naked eye at adulteration level ≤ 1 % despite the increase in the apparent protein content by about 46 % (or 43 % in the case of urea) at 1 % (w/w) dry blending of the chemical adulterants with the fresh samples. However, the spectral profiles of CO and PO FMP types highlighted that CO had high specificity for the

chemical adulterants while PO appeared to have contrasting abilities and masking their presence. The three multilevel SIMCA models built on the NIR spectral profiles of the typical and atypical FMP types using HMS as a tool were able to detect, confirm and differentiate the adulterations with an efficiency ranging from 89.8 to 100.0 %. The LOD of the CO, PO, SBO and SFO FMP models for melamine were 0.01 % (with specificity values ranging from 98.3 % to 100 %) and that for urea were 0.01 % (specificity 100 %), 0.01 % (specificity 100 %), 0.08 % (specificity 93.3 %) and 0.20 % (specificity 90.0 %) respectively considering the possible limits of adulteration levels defined for this study. Overall, all the models for the typical FMP types had LOD of at least 0.01 % for melamine and LOD ranging from 0.01 % to 0.20 % for urea adulterations. Thus, NIRS had lower LOD for melamine than urea in FMP. On the other hand, NIRS had lower LOD for urea in CO and PO FMP than in SBO and SFO FMP types. That implies that urea was more difficult to detect in FMP types containing vegetable oils of higher unsaturation in lipids. Similarly, the SIMCA models for the FMP types formulated with the unsaturated lipids (i.e., SBO and SFO) were less efficient in identifying and differentiating the melamine adulterated objects. The impact of the HMS screening approach was noteworthy as it helped in overcoming the effects of the complex sample matrices of the mixture of FMP and the adulterants by creating a sequence of high-performance SIMCA multilevel models with high interpretability. The PLSR models developed using the optimal wavelength intervals (i.e., iPLS) were able to predict the concentrations of the chemical adulterants in FMP with minimal errors ($R^2_p \geq 0.96$ and RSR ≤ 0.19) at 1.00 % adulteration level. The melamine adulterated FMP types were better predicted than those of urea as they mostly had lower RSR values. Overall, the PLSR models were robust enough for both calibrations and predictions, and the figures of merit met the required conditions to achieve an optimal generalisation. This study has demonstrated, for the first time, that NIRS is potentially useful for cost-effective detection and quantification of melamine and urea in FMP in order to protect the public health.

The accelerated storage of the FMP types for 7 weeks at 40 °C and 90 % RH significantly altered their FA, AA and NIR spectral profiles, although the packaging material utilised for the trials may likely differ from those used by large-scale commercial manufacturers. The stearic acid content significantly decreased while an apparent increase in leucine concentration was recorded for each of the milk fat replacements with vegetable oils under environmental stress. There was a statistically significant decrease in essential and non-essential AA but the percentage change in essential and conditionally essential AA decreased with increasing

unsaturation of the incorporated vegetable oils in FMP. Thus, the higher the unsaturation of the backbone fats, the higher the protective influence they may exert on the essential AA contents of FMP. The NIR spectra profiles of the FMP samples were generally altered upon storage with the peak intensity of the O-H functional group increasing due to water absorption and that of the C-H and C=O functional groups decreasing due to lipid-protein interactions. Aged CO and SFO FMP types appeared to have unique band alterations near 1685 nm and 1530 nm respectively while the aged samples containing the unsaturated vegetable oils may be related to the alterations near 1225 nm, 1725 nm and 2212 nm and those markers may be useful in discriminating them. Overall, the PO FMP type recorded the least spectral and FA profile alterations and as such appeared to be the most stable, and closest in stability to WMP, across the 7 weeks of storage. However, milk fat replacement with vegetable oils did not make FMP more stable than WMP. The multiclass SIMCA models, particularly those built with the combination of KS algorithm technique and the optimal wavelengths (i.e., 1722 – 1895 nm, 347 wavelengths), supported the PCA analysis in providing firm differentiation (i.e., zero false positive and 100 % specificity) of the fresh and the aged FMP samples, thus emphasising the magnitude of the compositional changes that may occur over time if the FMP products were subjected to such environmental stresses (i.e., 40 °C and 90 % RH) typical of the tropical countries. The results obtained in this study suggest that the KS algorithm sampling technique may enhance the classification modelling efficiency of FMP where samples were collected on weekly basis over time for the purpose of model development. The PLSR models developed using a combination of RS method and iPLS technique predicted the storage time with high efficiency (NSE ≥ 0.90) and low errors (RSR ≤ 0.28) indicative of FMP freshness and stability. The predictions supported value for variable selection and the result showed that the length of time for which the PO FMP type had been in storage under the tropical conditions may be predicted with a loss of accuracy, indicative of its high stability. Overall, this study has demonstrated the utility of NIRS as a quality assurance method to monitor changes in FMP during storage under inclement conditions.

The 7 FMP brands collected from Nigeria to validate the applicability of the qualitative and quantitative models developed in this project were found to be of acceptable process hygiene quality and microbiological safety. In addition, the Nigerian samples were most likely not adulterated with either melamine or urea since their SIMCA validations resulted in zero false positives (i.e., specificities = 100 %) and the prominent absorption peaks for these chemical

adulterants were not sighted in their NIR spectral profiles. This finding is concordant with the low level of microbial contaminants found in the FMP brands, although specific tests for melamine and urea were not undertaken on the samples due to cost limitations. The FA profile analysis supported the NIR spectral comparison/validation in revealing that PO formed a major component of the Nigerian products. Consequently, the 7 FMP brands were found to be less atherogenic (IA = 1.7) than WMP but contained relatively high LA to ALA ratios of about 27.8:1 to 63:1. Proline, leucine and lysine were the most abundant essential AA in the FMP samples but their protein contents appeared to be below (i.e., < 25.7 g/100 g of milk) the permitted limit. The deteriorative effects of the accelerated storage (i.e., at 40 °C and 90 % RH for 7 weeks) of the Nigerian FMP samples were reflected in their FA, AA and NIR spectral profiles, notwithstanding that the trials were conducted in their respective commercial packaging. In particular, the tryptophan and the lysine contents decreased upon storage due to lipid-protein co-oxidation and Maillard reactions defects respectively. Also, the peak NIRS absorbance of the O-H functional group increased with storage time, thus lowering the glass transition temperature and leading to lactose crystallization. The post storage extinction of the NIRS absorption bands at 1598 and 2100 nm confirmed the possible reaction of lactose and milk proteins whereas the sharp decay in band intensity at 2304 nm signified the progression of lipid oxidation process. Overall, most of the differential features observed in the quality parameters of the FMP brands during storage under inclement conditions resulted from the deteriorative effects of Maillard reaction, lipid-protein co-oxidation and lactose crystallisation. The multiclass SIMCA analyses performed in this study supported the PCA in showing that the fresh FMP brands were completely distinct from one another, and from those that were exposed to inclement conditions typical of the tropical countries. The PLSR models developed using only the iPLS optimal wavelength intervals predicted the storage time of the 7 FMP brands with high efficiency (NSE ≥ 0.74) and low errors (RSR≤ 0.46), with the Dano FMP having the best predictions (NSE = 1.00, RSR = 0.04) and the Luna brand the worst (NSE = 0.74, RSR = 0.46 (optimal < 0.50)). These results correlated with the barrier properties of Dano and Luna FMP brands, indicative of their freshness and stability. This study has demonstrated the practical relevance of our qualitative and quantitative models in the assessment of quality and safety of FMP products.

8.3 Main contributions of the research to body of knowledge

This research project has provided the protocols for rapid, inexpensive on-farm segregation of raw milk by herds for the manufacture of made-to-order high value niche products. In addition, the study appears to be the first to apply NIRS to FMP (a poorly studied product in high demand) and has provided chemical profiles and spectral analytical chemometric models necessary for the following:

- Comparison of the baseline variances in different formulations of FMP with respect to their FA and AA profiles, in addition to the nutritional consequences of the backbone fats used in their manufacture.

- Routine classification of FMP products for quality assurance monitoring.

- Efficient and cost-effective screening to detect, confirm, differentiate and quantify chemical adulterants in FMP products in order to reduce the threat to public health and safety of the many.

- Monitoring the stability and the compositional alterations in FMP during storage and distribution under inclement conditions.

8.4 Future research directions

Although this research has demonstrated that NIRS can reliably be used for precise, rapid and low-cost characterisation of the quality and safety of FMP (a commodity that has never before studied), the findings, conclusions and limitations reported from this study showed that the following gaps and opportunities exist to extend the methods for wider adoption.

- The four vegetable oils used in this study were independently utilised in our formulation but it will be interesting to investigate whether two or more oils that offer low LA/ALA ratio can be judiciously blended together in optimal proportions to formulate FMP with superior health benefits so as to be able to justify the fat inclusion on health grounds. In addition, we used freeze-drying method for our formulations and it would be pertinent to repeat these experiments using techniques like spray-drying that are more commercially widespread, to assess the possible impact of any variations.

- The minimum adulteration level (i.e., 0.01 %) defined for this study is higher than the threshold of 0.0001 % (1 ppm) set by Codex for IMP and a further study will be needed to test the efficacy of NIRS for low-level melamine and urea detection in FMP products.

- Additional studies of FMP types (i.e., containing CO, PO, SBO and SFO) stored under different humidity and temperature conditions using commercial packaging materials would be required for more extensive and accurate modelling.

- Alternative methods to PCA such as the Uniform Manifold Approximation and Projection technique for data structure analysis of the NIR spectra may be explored to improve dimension reduction quality and visualisation.

References

Abbas, O., Lecler, B., Dardenne, P., & Baeten, V. (2013). Detection of melamine and cyanuric acid in feed ingredients by near infrared spectroscopy and chemometrics. *Journal of Near Infrared Spectroscopy, 21*(3), 183-194.

Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley interdisciplinary reviews: computational statistics, 2*(1), 97-106.

Abernethy, D., Sheehan, C., Griffiths, J., & Williams, R. (2009). Adulteration of drugs and foods: compendial approaches to lowering risk. *Clinical Pharmacology & Therapeutics, 85*(4), 444-447.

Aernouts, B., Polshin, E., Lammertyn, J., & Saeys, W. (2011). Visible and near-infrared spectroscopic analysis of raw milk for cow health monitoring: Reflectance or transmittance? *Journal of dairy science, 94*(11), 5315-5329.

Afseth, N. K., & Kohler, A. (2012). Extended multiplicative signal correction in vibrational spectroscopy, a tutorial. *Chemometrics and Intelligent Laboratory Systems, 117*, 92-99.

Afseth, N. K., Segtnan, V. H., & Wold, J. P. (2006). Raman spectra of biological samples: A study of preprocessing methods. *Applied spectroscopy, 60*(12), 1358-1367.

Agelet, L. E., & Hurburgh, C. R., Jr. (2010). A tutorial on near infrared spectroscopy and its calibration. *Critical Reviews in Analytical Chemistry, 40*(4), 246-260.

Al-Qadiri, H., Lin, M., Al-Holy, M., Cavinato, A., & Rasco, B. A. (2008). Monitoring quality loss of pasteurized skim milk using visible and short wavelength near-infrared spectroscopy and multivariate analysis. *Journal of dairy science, 91*(3), 950-958.

Andersen, C. M., & Bro, R. (2010). Variable selection in regression—a tutorial. *Journal of Chemometrics, 24*(11-12), 728-737.

Andueza, D., Agabriel, C., Constant, I., Lucas, A., & Martin, B. (2013). Using visible or near infrared spectroscopy (NIRS) on cheese to authenticate cow feeding regimes. *Food chemistry, 141*(1), 209-214.

Angulo, A., Romera, J., Ramirez, M., & Gil, A. (1998). Effects of storage conditions on lipid oxidation in infant formulas based on several protein sources. *Journal of the American Oil Chemists' Society, 75*(11), 1603-1607.

ANZ Banking Group. (2018). New Zealand Dairy Update. Retrieved from https://www.anz.co.nz/content/dam/anzconz/documents/economics-and-market-research/2018/ANZ-New-Zealand-Dairy-Update-20180509.pdf?MOD=AJPERES

Azcarate, S. M., Gil, R., Smichowski, P., Savio, M., & Camiña, J. M. (2017). Chemometric application in foodomics: Nutritional quality parameters evaluation in milk-based infant formula. *Microchemical Journal, 130*, 1-6.

Balabin, R. M., & Smirnov, S. V. (2011). Melamine detection by mid-and near-infrared (MIR/NIR) spectroscopy: a quick and sensitive method for dairy products analysis including liquid milk, infant formula, and milk powder. *Talanta, 85*(1), 562-568.

Ballabio, D., & Todeschini, R. (2009). Multivariate classification for qualitative analysis. *Infrared spectroscopy for food quality analysis and control, 83*, e102.

Barbano, D., & Lynch, J. (2006). Major advances in testing of dairy products: Milk component and dairy product attribute testing. *Journal of dairy science, 89*(4), 1189-1194.

Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics: A Journal of the Chemometrics Society, 17*(3), 166-173.

Beć, K. B., & Huck, C. W. (2019). Breakthrough potential in near-infrared spectroscopy: Spectra simulation. A review of recent developments. *Frontiers in chemistry, 7*.

Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of chromatography A, 1158*(1-2), 196-214.

Borin, A., Ferrao, M. F., Mello, C., Maretto, D. A., & Poppi, R. J. (2006). Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk. *Analytica chimica acta, 579*(1), 25-32.

Botros, L. L., Jablonski, J., Chang, C., Bergana, M. M., Wehling, P., Harnly, J. M., . . . Moore, J. C. (2013). Exploring authentic skim and nonfat dry milk powder variance for the development of nontargeted adulterant detection methods using near-infrared spectroscopy and chemometrics. *Journal of agricultural and food chemistry, 61*(41), 9810-9818.

Boysworth, M. K., & Booksh, K. S. (2001). Aspects of multivariate calibration applied to near infrared spectroscopy. In D. A. Burns & E. W. Ciurczak (Eds.), *Handbook of Near-Infrared Analysis* (Second ed., pp. 209–240). New York: Marcel-Dekker Inc.

Boysworth, M. K., & Booksh, K. S. (2007). Aspects of Multivariate Calibration Applied to Near-Infrared Spectroscopy In D. A. Burns & E. W. Ciurczak (Eds.), *Handbook of Near-Infrared Analysis* (3 ed., pp. 207 - 228). New York: CRC Press.

Brereton, R. G. (2003). *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. Chichester, England: John Wiley & Sons Ltd.

Bruun, S. W., Søndergaard, I., & Jacobsen, S. (2007). Analysis of protein structures and interactions in complex food by near-infrared spectroscopy. 1. Gluten powder. *Journal of agricultural and food chemistry, 55*(18), 7234-7243.

Capuano, E., Boerrigter-Eenling, R., Koot, A., & van Ruth, S. M. (2015). Targeted and untargeted detection of skim milk powder adulteration by near-infrared spectroscopy. *Food analytical methods, 8*(8), 2125-2134.

Central Bank of Nigeria. (2016). Measuring Informal Cross-border Trade in Nigeria. Retrieved from https://www.cbn.gov.ng/out/2018/sd/measuring%20informal%20cross-border%20trade%20in%20nigeria.pdf

Chan, Y. (2003). Biostatistics 104: correlational analysis. *Singapore Med J, 44*(12), 614-619.

Chen, H., Tan, C., Lin, Z., & Wu, T. (2017). Detection of melamine adulteration in milk by near-infrared spectroscopy and one-class partial least squares. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 173*, 832-836.

Chen, H., Tan, C., Lin, Z., & Wu, T. (2018). Classification and quantitation of milk powder by near-infrared spectroscopy and mutual information-based variable selection and partial least squares. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 189*, 183-189.

Cheng, H., Zhu, R.-G., Erichsen, H., Soerensen, J., Petersen, M. A., & Skibsted, L. H. (2017). High temperature storage of infant formula milk powder for prediction of storage stability at ambient conditions. *International Dairy Journal, 73*, 166-174.

Cheng, Y., Dong, Y., Wu, J., Yang, X., Bai, H., Zheng, H., . . . Li, M. (2010). Screening melamine adulterant in milk powder with laser Raman spectrometry. *Journal of Food Composition and Analysis, 23*(2), 199-202.

Cho, Y.-J., & Kang, S. (2011). *Emerging technologies for food quality and food safety evaluation* (1st Edition ed.). Boca Raton: CRC Press.

Chong, I.-G., & Jun, C.-H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems, 78*(1-2), 103-112.

Chopovda, V., Clarke, R., Fowler, A., Fullard, L., Goodman, J., Hall, L., & Taylor, S. W. (2016). Predicting the shelf life of milk powder. *ANZIAM Journal, 58*, 379-421.

Christy, A. A., Kasemsumran, S., Du, Y., & Ozaki, Y. (2004). The detection and quantification of adulteration in olive oil by near-infrared spectroscopy and chemometrics. *Analytical Sciences, 20*(6), 935-940.

CLAL-Dairy Economic Consulting firm. (2019). EU-28: Dairy sector. Retrieved from https://www.clal.it/en/?section=stat_ue15

Codex Alimentarius Commission. (1999). Elaboration of a standard for products in which milk components are substituted by non-milk components Retrieved from http://www.fao.org/tempref/codex/Meetings/CCASIA/ccasia12/as99_06e.pdf

Codex Alimentarius Commission. (2006). Standard for a blend of skimmed milk and vegetable fat in powdered form. Codex STAN 251-2006.

Codex Alimentarius Commission. (2013). General standard for contaminants and toxins in food and feed. (CODEX STAN 193-1995).

Collomb, M., Bütikofer, U., Sieber, R., Jeangros, B., & Bosset, J.-O. (2002). Composition of fatty acids in cow's milk fat produced in the lowlands, mountains and highlands of Switzerland using high-resolution gas chromatography. *International Dairy Journal, 12*(8), 649-659.

Coppa, M., Ferlay, A., Leroux, C., Jestin, M., Chilliard, Y., Martin, B., & Andueza, D. (2010). Prediction of milk fatty acid composition by near infrared reflectance spectroscopy. *International Dairy Journal, 20*(3), 182-189.

Coppa, M., Martin, B., Agabriel, C., Chassaing, C., Sibra, C., Constant, I., . . . Andueza, D. (2012). Authentication of cow feeding and geographic origin on milk using visible and near-infrared spectroscopy. *Journal of dairy science, 95*(10), 5544-5551.

Craig, A. P., Botelho, B. G., Oliveira, L. S., & Franca, A. S. (2018). Mid infrared spectroscopy and chemometrics as tools for the classification of roasted coffees by cup quality. *Food chemistry, 245*, 1052-1061.

Crapiste, G. H., Brevedan, M. I., & Carelli, A. A. (1999). Oxidation of sunflower oil during storage. *Journal of the American Oil Chemists' Society, 76*(12), 1437.

Davies, M. J. (2005). The oxidative environment and protein damage. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 1703*(2), 93-109.

De la Roza-Delgado, B., Garrido-Varo, A., Soldado, A., Arrojo, A. G., Valdés, M. C., Maroto, F., & Pérez-Marín, D. (2017). Matching portable NIRS instruments for in situ monitoring indicators of milk composition. *Food Control, 76*, 74-81.

De Lorgeril, M., & Salen, P. (2004). Alpha-linolenic acid and coronary heart disease. *Nutrition, Metabolism and Cardiovascular Diseases, 14*(3), 162-169.

De Souza Gondim, C., Junqueira, R. G., De Souza, S. V. C., Ruisánchez, I., & Callao, M. P. (2017). Detection of several common adulterants in raw milk by MID-infrared spectroscopy and one-class and multi-class multivariate strategies. *Food chemistry, 230*, 68-75.

Dierking, R., Kallenbach, R., & Roberts, C. (2010). Fatty acid profiles of orchardgrass, tall fescue, perennial ryegrass, and alfalfa. *Crop Science, 50*(1), 391-402.

Ding, H., & Xu, R. (2000). Near-infrared spectroscopic technique for detection of beef hamburger adulteration. *Journal of agricultural and food chemistry, 48*(6), 2193-2198.

Dooley, A., Parker, W., Blair, H., & Hurley, E. (2005). Implications of on-farm segregation for valuable milk characteristics. *Agricultural Systems, 85*(1), 82-97.

Early, R. (1988). *Technology of dairy products* (Second ed.). London: Blackie Academic & Professional.

Early, R. (1998). *Technology of dairy products*: Springer Science & Business Media.

Ejeahalaka, K. K., & On, S. L. (2019a). Chemometric studies of the effects of milk fat replacement with different proportions of vegetable oils in the formulation of fat-filled milk powders: Implications for quality assurance. *Food chemistry, 295*, 198-205.

Ejeahalaka, K. K., & On, S. L. (2019b). Effective detection and quantification of chemical adulterants in model fat-filled milk powders using NIRS and hierarchical modelling strategies. *Food chemistry*, 125785.

Ejeahalaka, K. K., & On, S. L. (2020). Characterisation of the quality alterations in model fat-filled milk powders under inclement conditions and the prediction of the storage time using near infrared spectroscopy. *Food chemistry*, 126752.

Elgersma, A. (2015). Grazing increases the unsaturated fatty acid concentration of milk from grass-fed cows: A review of the contributing factors, challenges and future perspectives. *European journal of lipid science and technology, 117*(9), 1345-1369.

Eludoyin, O. M., Adelekan, I. O., Webster, R., & Eludoyin, A. O. (2014). Air temperature, relative humidity, climate regionalization and thermal comfort of Nigeria. *International Journal of Climatology, 34*(6), 2000-2018.

Estévez, M., Kylli, P., Puolanne, E., Kivikari, R., & Heinonen, M. (2008). Oxidation of skeletal muscle myofibrillar proteins in oil-in-water emulsions: interaction with lipids and effect of selected phenolic compounds. *Journal of agricultural and food chemistry, 56*(22), 10933-10940.

Estévez, M., & Xiong, Y. (2019). Intake of oxidized proteins and amino acids and causative oxidative stress and disease: Recent scientific evidences and hypotheses. *Journal of food science, 84*(3), 387-396.

European Commission. (2017). Milk Market Observatory. Retrieved from https://ec.europa.eu/info/sites/info/files/food-farming-fisheries/farming/documents/mmo-report-2017-09-26_en.pdf
Everstine, K., Spink, J., & Kennedy, S. (2013). Economically motivated adulteration (EMA) of food: common characteristics of EMA incidents. *Journal of food protection, 76*(4), 723-735.
FAO. (2010). International experts limit Melamine levels in food. Retrieved from http://www.fao.org/news/story/en/item/43719/icode/
FAO/WHO. (2003). Assuring Food Safety and Quality: Guidelines for Strengthening National Food Control Systems. Retrieved from http://www.fao.org/3/y8705e/y8705e00.htm
Fievez, V., Vlaeminck, B., Jenkins, T., Enjalbert, F., & Doreau, M. (2007). Assessing rumen biohydrogenation and its manipulation in vivo, in vitro and in situ. *European journal of lipid science and technology, 109*(8), 740-756.
Finete, V. d. L. M., Gouvêa, M. M., de Carvalho Marques, F. F., & Netto, A. D. P. (2013). Is it possible to screen for milk or whey protein adulteration with melamine, urea and ammonium sulphate, combining Kjeldahl and classical spectrophotometric methods? *Food chemistry, 141*(4), 3649-3655.
Fleming, A., Schenkel, F., Chen, J., Malchiodi, F., Bonfatti, V., Ali, R., . . . Miglior, F. (2017). Prediction of milk fatty acid content with mid-infrared spectroscopy in Canadian dairy cattle using differently distributed model development sets. *Journal of dairy science, 100*(6), 5073-5081.
Ford, J., Hurrell, R., & Finot, P. (1983). Storage of milk powders under adverse conditions: 2. Influence on the content of water-soluble vitamins. *British Journal of Nutrition, 49*(3), 355-364.
Frankel, E. N. (1984). Lipid oxidation: mechanisms, products and biological significance. *Journal of the American Oil Chemists' Society, 61*(12), 1908-1917.
Frankhuizen, R. (2001). NIR analysis of dairy products. In E. W. Ciurczak & D. A. Burns (Eds.), *Handbook of near-infrared analysis* (pp. 499-535). New York: CRC Press.
García-Sánchez, F., Galvez-Sola, L., Martínez- Nicolás, J. J., Muelas-Domingo, R., & Nieves, M. (2017). Using Near-Infrared Spectroscopy in Agricultural Systems. In K. Kyprianidis & J. Skvaril (Eds.), *Developments in Near-Infrared Spectroscopy* (pp. 97-127). London: IntechOpen.
Gibson, R. A., Makrides, M., Neumann, M. A., Simmer, K., Mantzioris, E., & James, M. J. (1994). Ratios of linoleic acid to α-linolenic acid in formulas for term infants. *The Journal of pediatrics, 125*(5), S48-S55.
Gordon, J. (1997). Dairy products. In M. Ranken, R. Kill, & C. Baker (Eds.), *Food industries manual* (24 ed., pp. 74-131). London: Blackie Academic & Professional.
Grewal, M. K., Chandrapala, J., Donkor, O., Apostolopoulos, V., Stojanovska, L., & Vasiljevic, T. (2017). Fourier transform infrared spectroscopy analysis of physicochemical changes in UHT milk during accelerated storage. *International Dairy Journal, 66*, 99-107.
Grummer, R. R. (1991). Effect of feed on the composition of milk fat. *Journal of dairy science, 74*(9), 3244-3257.
Gunstone, F. D. (2011). Production and trade of vegetable oils. In F. D. Gunstone (Ed.), *Vegetable Oils in Food Technology: Composition, Properties and Uses* (Second ed., pp. 1-17).
Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research, 3*(Mar), 1157-1182.
Handford, C. E., Campbell, K., & Elliott, C. T. (2016). Impacts of milk fraud on food safety and nutrition with special emphasis on developing countries. *Comprehensive Reviews in Food Science and Food Safety, 15*(1), 130-142.
Hansen, P. S. (1980). Production of agglomerated fat-filled milk powder. *International Journal of Dairy Technology, 33*(1), 19-23.
Hanuš, O., Samková, E., Křížová, L., Hasoňová, L., & Kala, R. (2018). Role of fatty acids in milk fat and the influence of selected factors on their variability—a review. *Molecules, 23*(7), 1636.
Heinrichs, J., Jones, C., & Bailey, K. (1997). Milk components: Understanding the causes and importance of milk fat and protein variation in your dairy herd. *Dairy Anim. Sci, 5*, 1e-8e.
Higgs, K., & Boland, M. J. (2014). Changes in milk proteins during storage of dry powders In A. Thompson, M. Boland, & H. Singh (Eds.), *Milk Proteins: from Expression to Food* (2nd Edition

ed., pp. 343-357). Retrieved from https://www-sciencedirect-com.ezproxy.lincoln.ac.nz/book/9780124051713/milk-proteins
Hilding-Ohlsson, A., Fauerbach, J. A., Sacco, N. J., Bonetto, M. C., & Cortón, E. (2012). Voltamperometric discrimination of urea and melamine adulterated skimmed milk powder. *Sensors, 12*(9), 12220-12234.
Holman, R. T., & Edmondson, P. R. (1956). Near-infrared spectra of fatty acids and some related substances. *Analytical Chemistry, 28*(10), 1533-1538.
Holman, R. T., Nickell, C., Privett, O. S., & Edmondson, P. R. (1958). Detection and measurement of hydroperoxides by near infrared spectrophotometry. *Journal of the American Oil Chemists' Society, 35*(8), 422-425.
Huang, J., Romero-Torres, S., & Moshgbar, M. (2010). Practical Considerations in Data Pre-treatment for NIR and Raman Spectroscopy. *American Pharmaceutical Review.*
Hubschmann, P. (2013). Arla Food - Dano Cool Cow Affordable Milk: A Nigerian Adventure. Retrieved from https://www.arlafoodsingredients.com/globalassets/global/about-us/csr/food-aid/peer-hubschmann-food-aid-120413.pdf
Huppertz, T., & Gazi, I. (2016). Lactose in dairy ingredients: Effect on processing and storage stability. *Journal of dairy science, 99*(8), 6842-6851.
Hurrell, R., Finot, P., & Ford, J. (1983). Storage of milk powders under adverse conditions: 1. Losses of lysine and of other essential amino acids as determined by chemical and microbiological methods. *British Journal of Nutrition, 49*(3), 343-354.
Hurtaud, C., Dutreuil, M., Coppa, M., Agabriel, C., & Martin, B. (2014). Characterization of milk from feeding systems based on herbage or corn silage with or without flaxseed and authentication through fatty acid profile. *Dairy Science & Technology, 94*(2), 103-123.
Ibrahim, N. A. (2011). The luaric (coconut and palm kernel) oils. In F. D. Gunstone (Ed.), *Vegetable oils in food technology: composition, properties and uses* (pp. 169-194).
ISO. (2002). Microbiology of food and animal feeding stuffs — Horizontal method for the detection of Salmonella spp. (ISO 6579) Retrieved from https://www.iso.org/standard/29315.html
ISO. (2003). Microbiology of food and animal feeding stuffs — Horizontal method for the enumeration of sulfite-reducing bacteria growing under anaerobic conditions (ISO 15213). Retrieved from https://www.iso.org/standard/26852.html
ISO. (2013). Microbiology of the food chain — Horizontal method for the enumeration of microorganisms — Part 2: Colony count at 30 degrees C by the surface plating technique (ISO 4833-2). Retrieved from https://www.iso.org/standard/59509.html
ISO/IDF. (2006). Milk and milk products - Detection of Enterobacter sakazakii (ISO 22964/IDF 210). Retrieved from https://www.iso.org/standard/41258.html
Jenkins, T., Wallace, R., Moate, P., & Mosley, E. (2008). Board-invited review: Recent advances in biohydrogenation of unsaturated fatty acids within the rumen microbial ecosystem. *Journal of animal science, 86*(2), 397-412.
Jensen, G. K., & Nielsen, P. (1982). Milk powder and recombination of milk and milk products. *Journal of dairy research, 49*(3), 515-544.
Jeon, I., Roberts, H., & Senecal, A. (1992). *Replacement of Coconut Oils with Unsaturated Oils in Recombined Filled Milk*. Retrieved from
Kaddour, A. A., Grand, E., Barouh, N., Baréa, B., Villeneuve, P., & Cuq, B. (2006). Near-infrared spectroscopy for the determination of lipid oxidation in cereal food products. *European journal of lipid science and technology, 108*(12), 1037-1046.
Kambhampati, S., Li, J., Evans, B. S., & Allen, D. K. (2019). Accurate and efficient amino acid analysis for protein quantification using hydrophilic interaction chromatography coupled tandem mass spectrometry. *Plant methods, 15*(1), 46.
Karoui, R., Mouazen, A. M., Dufour, É., Pillonel, L., Picque, D., Bosset, J.-O., & De Baerdemaeker, J. (2006). Mid-infrared spectrometry: A tool for the determination of chemical parameters in Emmental cheeses produced during winter. *Le Lait, 86*(1), 83-97.
Kasemsumran, S., Thanapase, W., & Kiatsoonthon, A. (2007). Feasibility of near-infrared spectroscopy to detect and to quantify adulterants in cow milk. *Analytical Sciences, 23*(7), 907-910.

Katz, G., Merin, U., Bezman, D., Lavie, S., Lemberskiy-Kuzin, L., & Leitner, G. (2016). Real-time evaluation of individual cow milk for higher cheese-milk quality with increased cheese yield. *Journal of dairy science, 99*(6), 4178-4187.

Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics, 11*(1), 137-148.

Keys, A., Mienotti, A., Karvonen, M. J., Aravanis, C., Blackburn, H., Buzina, R., . . . Keys, M. H. (1986). The diet and 15-year death rate in the seven countries study. *American journal of epidemiology, 124*(6), 903-915.

Koh, G., Chia, R. S., Lin, Q., Cheow, P. S., Teo, T. L., & Lee, T. K. (2011). Determination of melamine in milk powder using gas chromatography–high-resolution isotope dilution mass spectrometry. *Journal of separation science, 34*(21), 3043-3052.

Kong, W., Zhang, C., Liu, F., Nie, P., & He, Y. (2013). Rice seed cultivar identification using near-infrared hyperspectral imaging and multivariate data analysis. *Sensors, 13*(7), 8916-8927.

Lasch, P. (2012). Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging. *Chemometrics and Intelligent Laboratory Systems, 117*, 100-114.

Lavine, B., & Workman, J. J. (2004). Chemometrics. *Analytical Chemistry, 76*(12), 3365-3371.

Lee, D.-S., Noh, B.-S., Bae, S.-Y., & Kim, K. (1998). Characterization of fatty acids composition in vegetable oils by gas chromatography and chemometrics. *Analytica chimica acta, 358*(2), 163-175.

Lindmark Månsson, H. (2008). Fatty acids in bovine milk fat. *Food & nutrition research, 52*(1), 1821.

Lloyd, M., Hess, S., & Drake, M. (2009). Effect of nitrogen flushing and storage temperature on flavor and shelf-life of whole milk powder. *Journal of dairy science, 92*(6), 2409-2422.

Lock, A. L., & Bauman, D. E. (2004). Modifying milk fat composition of dairy cows to enhance fatty acids beneficial to human health. *Lipids, 39*(12), 1197-1206.

Lu, C., Xiang, B., Hao, G., Xu, J., Wang, Z., & Chen, C. (2009). Rapid detection of melamine in milk powder by near infrared spectroscopy. *Journal of Near Infrared Spectroscopy, 17*(2), 59-67.

Ma, Y., Dong, W., Bao, H., Fan, C., Fang, Y., & Wang, S. (2016). Kinetic Determination of Urea in Milk Powder by Nonlinear Chemical Fingerprint Technique. *Food and Nutrition Sciences, 7*(7), 495-503.

Manning, L., & Soon, J. M. (2016). Food safety, food fraud, and food defense: a fast evolving literature. *Journal of food science, 81*(4), R823-R834.

Marchitelli, C., Contarini, G., De Matteis, G., Crisà, A., Pariset, L., Scatà, M. C., . . . Moioli, B. (2013). Milk fatty acid variability: effect of some candidate genes involved in lipid synthesis. *Journal of dairy research, 80*(2), 165-173.

Martens, H., & Stark, E. (1991). Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. *Journal of pharmaceutical and biomedical analysis, 9*(8), 625-635.

Martin, B., Fedele, V., Ferlay, A., Grolier, P., Rock, E., Gruffat, D., & Chilliard, Y. (2004). *Effects of grass-based diets on the content of micronutrients and fatty acids in bovine and caprine dairy products.* Paper presented at the Land use systems in grassland dominated regions. Proceedings of the 20th General Meeting of the European Grassland Federation, Luzern, Switzerland, 21-24 June 2004.

Mauer, L. J., Chernyshova, A. A., Hiatt, A., Deering, A., & Davis, R. (2009). Melamine detection in infant formula powder using near-and mid-infrared spectroscopy. *Journal of agricultural and food chemistry, 57*(10), 3974-3980.

Meilgaard, M. C., Carr, B. T., & Civille, G. V. (1999). *Sensory evaluation techniques*: CRC press.

Metrohm AG. (2014). A guide to near-infrared spectroscopic analysis of industrial manufacturing processes. Retrieved from https://www.metrohm.com/en/documents/81085026

Minitab 18 Statistical Software. (2017). Computer software. Retrieved from www.minitab.com

Modler, H. W., Rippen, A., & Stine, C. (1970). Physical and chemical stability of soybean oil-filled milk. *Journal of food science, 35*(3), 302-305.

Moore, J. C., DeVries, J. W., Lipp, M., Griffiths, J. C., & Abernethy, D. R. (2010). Total protein methods and their potential utility to reduce the risk of food protein adulteration. *Comprehensive Reviews in Food Science and Food Safety, 9*(4), 330-357.

Moore, J. C., Spink, J., & Lipp, M. (2012). Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. *Journal of food science, 77*(4), R118-R126.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE, 50*(3), 885-900.

Mouazen, A., Dridi, S., Rouissi, H., De Baerdemaeker, J., & Ramon, H. (2009). Prediction of selected ewe's milk properties and differentiating between pasture and box feeding using visible and near infrared spectroscopy. *Biosystems engineering, 104*(3), 353-361.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology, 10*(3), 282-290.

Nawrocka, A., & Lamorska, J. (2013). Determination of food quality by using spectroscopic methods. In *Advances in agrophysical research*: IntechOpen.

New Zealand Milk Products. (2019). Product Summary: Fat Filled Milk Powder. Retrieved from https://www.nzmp.com/content/dam/nzmp/pdfs/nzmp-fat-filled-milk-powder-product-summary.pdf

Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J. P., Munck, L., & Engelsen, S. B. (2000). Interval Partial Least-Squares Regression (i PLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. *Applied spectroscopy, 54*(3), 413-419.

Ntakatsane, M., Liu, X., & Zhou, P. (2013). Rapid detection of milk fat adulteration with vegetable oil by fluorescence spectroscopy. *Journal of dairy science, 96*(4), 2130-2136.

Núñez-Sánchez, N., Martínez-Marín, A., Polvillo, O., Fernández-Cabanás, V., Carrizosa, J., Urrutia, B., & Serradilla, J. (2016). Near infrared spectroscopy (NIRS) for the determination of the milk fat fatty acid profile of goats. *Food chemistry, 190*, 244-252.

Oftedal, O. T., Eisert, R., & Barrell, G. K. (2014). Comparison of analytical and predictive methods for water, protein, fat, sugar, and gross energy in marine mammal milk. *Journal of dairy science, 97*(8), 4713-4732.

Oliveri, P. (2017). Class-modelling in food analytical chemistry: development, sampling, optimisation and validation issues–a tutorial. *Analytica chimica acta, 982*, 9-19.

Oliveri, P., Di Egidio, V., Woodcock, T., & Downey, G. (2011). Application of class-modelling techniques to near infrared data for food authentication purposes. *Food chemistry, 125*(4), 1450-1456.

Oliveri, P., & Downey, G. (2012). Multivariate class modeling for the verification of food-authenticity claims. *Trends in Analytical Chemistry, 35*, 74-86.

Osborne, B. G. (2006). Near-infrared spectroscopy in food analysis. *Encyclopedia of analytical chemistry: applications, theory and instrumentation*.

Osborne, B. G., Fearn, T., & Hindle, P. H. (1993). *Practical NIR spectroscopy with applications in food and beverage analysis* (Second ed.). London: Longman Scientific & Technical.

Pacheco, Y. M., Bermúdez, B., López, S., Abia, R., Villar, J., & Muriana, F. J. (2006). Ratio of oleic to palmitic acid is a dietary determinant of thrombogenic and fibrinolytic factors during the postprandial state in men–. *The American journal of clinical nutrition, 84*(2), 342-349.

Pasquini, C. (2003). Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *Journal of the Brazilian chemical society, 14*(2), 198-219.

Pei, X., Tandon, A., Alldrick, A., Giorgi, L., Huang, W., & Yang, R. (2011). The China melamine milk scandal and its implications for food safety regulation. *Food policy, 36*(3), 412-420.

Pérez-Enciso, M., & Tenenhaus, M. (2003). Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. *Human genetics, 112*(5-6), 581-592.

Pomerantsev, A. L., & Rodionova, O. Y. (2018). Multiclass partial least squares discriminant analysis: Taking the right way—A critical tutorial. *Journal of Chemometrics, 32*(8), e3030.

Poonia, A., Jha, A., Sharma, R., Singh, H. B., Rai, A. K., & Sharma, N. (2017). Detection of adulteration in milk: A review. *International Journal of Dairy Technology, 70*(1), 23-42.
Porep, J. U., Kammerer, D. R., & Carle, R. (2015). On-line application of near infrared (NIR) spectroscopy in food production. *Trends in Food Science & Technology, 46*(2), 211-230.
Powers, D. M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies, 2*(1), 37-63.
Prosser, C., Svetashev, V., Vyssotski, M., & Lowry, D. (2010). Composition and distribution of fatty acids in triglycerides from goat infant formulas with milk fat. *Journal of dairy science, 93*(7), 2857-2862.
Qin, J., Chao, K., & Kim, M. S. (2013). Simultaneous detection of multiple adulterants in dry milk using macro-scale Raman chemical imaging. *Food chemistry, 138*(2-3), 998-1007.
R Core Team. (2018). R: A language and environment for statistical computing. Retrieved from http://www.R-project.org
Reis, N., Botelho, B. G., Franca, A. S., & Oliveira, L. S. (2017). Simultaneous detection of multiple adulterants in ground roasted coffee by ATR-FTIR spectroscopy and data fusion. *Food analytical methods, 10*(8), 2700-2709.
Renner, E. (1988). Storage stability and some nutritional aspects of milk powders and ultra high temperature products at high ambient temperatures. *Journal of dairy research, 55*(1), 125-142.
Richards, E. L., & Chandrasekhara, M. (1960). Chemical changes in dried skim-milk during storage. *Journal of dairy research, 27*(1), 59-66.
Rinnan, Å., Van Den Berg, F., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry, 28*(10), 1201-1222.
Robert, P., Bertrand, D., Devaux, M. F., & Grappin, R. (1987). Multivariate analysis applied to near-infrared spectra of milk. *Analytical Chemistry, 59*(17), 2187-2191.
Rodríguez-Alcalá, L. M., García-Martínez, M. C., Cachón, F., Marmesat, S., Alonso, L., Márquez-Ruiz, G., & Fontecha, J. (2007). Changes in the lipid composition of powdered infant formulas during long-term storage. *Journal of agricultural and food chemistry, 55*(16), 6533-6538.
Roggo, Y., Duponchel, L., & Huvenne, J.-P. (2003). Comparison of supervised pattern recognition methods with McNemar's statistical test: Application to qualitative analysis of sugar beet by near-infrared spectroscopy. *Analytica chimica acta, 477*(2), 187-200.
Romeu-Nadal, M., Chavez-Servin, J., Castellote, A., Rivero, M., & Lopez-Sabater, M. (2007). Oxidation stability of the lipid fraction in milk powder formulas. *Food chemistry, 100*(2), 756-763.
Rugoho, I., Liu, Y., & Dewhurst, R. (2014). Analysis of major fatty acids in milk produced from high-quality grazed pasture. *New Zealand journal of agricultural research, 57*(3), 165-179.
Santos, P., Pereira-Filho, E., & Rodriguez-Saona, L. (2013). Rapid detection and quantification of milk adulteration using infrared microspectroscopy and chemometrics analysis. *Food chemistry, 138*(1), 19-24.
Saptoro, A., Tadé, M. O., & Vuthaluru, H. (2012). A modified Kennard-Stone algorithm for optimal division of data for developing artificial neural network models. *Chemical Product and Process Modeling, 7*(1).
Šašić, S., & Ozaki, Y. (2001a). Short-wave near-infrared spectroscopy of biological fluids. 1. Quantitative analysis of fat, protein, and lactose in raw milk by partial least-squares regression and band assignment. *Analytical Chemistry, 73*(1), 64-71.
Šašić, S., & Ozaki, Y. (2001b). Wavelength—Wavelength and Sample—Sample Two-Dimensional Correlation Analyses of Short-Wave Near-Infrared Spectra of Raw Milk. *Applied spectroscopy, 55*(2), 163-172.
Savitzky, A., & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry, 36*(8), 1627-1639.
Scheidegger, D., Radici, P. M., Vergara-Roig, V. A., Bosio, N. S., Pesce, S. F., Pecora, R. P., . . . Kivatinitz, S. C. (2013). Evaluation of milk powder quality by protein oxidative modifications. *Journal of dairy science, 96*(6), 3414-3423.

Schmid, M., Dallmann, K., Bugnicourt, E., Cordoni, D., Wild, F., Lazzeri, A., & Noller, K. (2012). Properties of whey-protein-coated films and laminates as novel recyclable food packaging materials with excellent barrier properties. *International Journal of Polymer Science, 2012*.

Schmidmeier, C., O'Gorman, C., Drapala, K., Waldron, D., & O'Mahony, J. (2019). Elucidation of factors responsible for formation of white flecks in reconstituted fat filled milk powders. *Colloids and Surfaces A: Physicochemical and Engineering Aspects, 575*, 245-255.

Schoder, D. (2010). Melamine milk powder and infant formula sold in East Africa. *Journal of food protection, 73*(9), 1709-1714.

Scholl, P. F., Bergana, M. M., Yakes, B. J., Xie, Z., Zbylut, S., Downey, G., . . . Holroyd, S. E. (2017). Effects of the adulteration technique on the near-infrared detection of melamine in milk powder. *Journal of agricultural and food chemistry, 65*(28), 5799-5809.

Shetty, N., Rinnan, Å., & Gislum, R. (2012). Selection of representative calibration sample sets for near-infrared reflectance spectroscopy to predict nitrogen concentration in grasses. *Chemometrics and Intelligent Laboratory Systems, 111*(1), 59-65.

Shingfield, K. J., Bonnet, M., & Scollan, N. D. (2013). Recent developments in altering the fatty acid composition of ruminant-derived foods. *Animal, 7*(s1), 132-162.

Siew, W. L. (2011). Palm oil. In F. D. Gunstone (Ed.), *Vegetable oils in food technology: composition, properties and uses* (pp. 25-54). Retrieved from https://onlinelibrary.wiley.com/doi/book/10.1002/9781444339925

Simopoulos, A. P. (2002). The importance of the ratio of omega-6/omega-3 essential fatty acids. *Biomedicine & pharmacotherapy, 56*(8), 365-379.

Sinelli, N., Limbo, S., Torri, L., Di Egidio, V., & Casiraghi, E. (2010). Evaluation of freshness decay of minced beef stored in high-oxygen modified atmosphere packaged at different temperatures using NIR and MIR spectroscopy. *Meat science, 86*(3), 748-752.

Skogholt, J., Liland, K. H., & Indahl, U. G. (2019). Preprocessing of spectral data in the extended multiplicative signal correction framework using multiple reference spectra. *Journal of Raman Spectroscopy, 50*(3), 407-417.

Smith, R., Inomata, H., & Peters, C. (2013). *Introduction to supercritical fluids: A spreadsheet-based approach* (Vol. 4). Oxford: Elsevier.

Sørensen, I., Neve, T., Ottosen, N., Larsen, L. B., Dalsgaard, T. K., & Wiking, L. (2017). Storage stability of whole milk powder produced from raw milk reverse osmosis retentate. *Dairy Science & Technology, 96*(6), 873-886.

Spink, J., & Moyer, D. C. (2011). Defining the public health threat of food fraud. *Journal of food science, 76*(9), R157-R163.

Stapelfeldt, H., Nielsen, B. R., & Skibsted, L. H. (1997). Effect of heat treatment, water activity and storage temperature on the oxidative stability of whole milk powder. *International Dairy Journal, 7*(5), 331-339.

Stark, E. W., & Martens, H. (1996). Multiplicative signal correction method and apparatus. In: Google Patents.

Stevens, A., & Ramirez-Lopez, L. (2014). An introduction to the prospectr package. *R Package Vignette, Report No.: R Package Version 0.1, 3*.

Suhr, D. D. (2005). *Principal component analysis vs. exploratory factor analysis (paper 203-30).* Paper presented at the Proceedings of the thirtieth annual SAS® users group international conference.

Sun, Y.-E., Wang, W.-D., Chen, H.-W., & Li, C. (2011). Autoxidation of unsaturated lipids in food emulsion. *Critical reviews in food science and nutrition, 51*(5), 453-466.

Talavera, L. (2005). *An evaluation of filter and wrapper methods for feature selection in categorical clustering.* Paper presented at the International Symposium on Intelligent Data Analysis.

Tham, T. W. Y., Xu, X., Yeoh, A. T. H., & Zhou, W. (2017). Investigation of caking by fat bridging in aged infant formula. *Food chemistry, 218*, 30-39.

Thomas, M. E., Scher, J., Desobry-Banon, S., & Desobry, S. (2004). Milk powders ageing: effect on physical and functional properties. *Critical reviews in food science and nutrition, 44*(5), 297-322.

Tsenkova, R., Atanassova, S., Itoh, K., Ozaki, Y., & Toyoda, K. (2000). Near infrared spectroscopy for biomonitoring: cow milk composition measurement in a spectral region from 1,100 to 2,400 nanometers. *Journal of animal science, 78*(3), 515-522.

Ulbricht, T., & Southgate, D. (1991). Coronary heart disease: seven dietary factors. *The lancet, 338*(8773), 985-992.

Uppu, P. (2001). *Determination of Packaging Material Requirements for Optimum Shelf Life of Packaged Filled Milk Powder.* Victoria University,

USDA. (2012). United States Department of Agriculture Foreign Agricultural Service - Exporter Guide (2012). Retrieved from https://apps.fas.usda.gov/newgainapi/api/report/downloadreportbyfilename?filename=Exporter%20Guide_Lagos_Nigeria_11-20-2012.pdf

Valenti, B., Martin, B., Andueza, D., Leroux, C., Labonne, C., Lahalle, F., . . . Brochard, M. (2013). Infrared spectroscopic methods for the discrimination of cows' milk according to the feeding system, cow breed and altitude of the dairy farm. *International Dairy Journal, 32*(1), 26-32.

Vargas-Bello-Pérez, E., Toro-Mujica, P., Enriquez-Hidalgo, D., Fellenberg, M. A., & Gómez-Cortés, P. (2017). Discrimination between retail bovine milks with different fat contents using chemometrics and fatty acid profiling. *Journal of dairy science, 100*(6), 4253-4257.

Vignolles, M.-L., Jeantet, R., Lopez, C., & Schuck, P. (2007). Free fat, surface fat and dairy powders: interactions between process and product. A review. *Le Lait, 87*(3), 187-236.

Vongsvivut, J., Heraud, P., Zhang, W., Kralovec, J. A., McNaughton, D., & Barrow, C. J. (2012). Quantitative determination of fatty acid compositions in micro-encapsulated fish-oil supplements using Fourier transform infrared (FTIR) spectroscopy. *Food chemistry, 135*(2), 603-609.

Wazir, H., Chay, S. Y., Zarei, M., Hussin, F. S., Mustapha, N. A., Ibadullah, W., . . . Saari, N. (2019). Effects of Storage Time and Temperature on Lipid Oxidation and Protein Co-Oxidation of Low-Moisture Shredded Meat Products. *Antioxidants, 8*(10), 486.

Westad, F., Schmidt, A., & Kermit, M. (2008). Incorporating chemical band-assignment in near infrared spectroscopy regression models. *Journal of Near Infrared Spectroscopy, 16*(3), 265-273.

Wishart, D. S. (2007). Current progress in computational metabolomics. *Briefings in bioinformatics, 8*(5), 279-293.

Wójcicki, K., Khmelinskii, I., Sikorski, M., & Sikorska, E. (2015). Near and mid infrared spectroscopy and multivariate data analysis in studies of oxidation of edible oils. *Food chemistry, 187*, 416-423.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. *Multivariate analysis*, 391-420.

Wold, J. P., Jakobsen, T., & Krane, L. (1996). Atlantic salmon average fat content estimated by near-infrared transmittance spectroscopy. *Journal of food science, 61*(1), 74-77.

Wold, S. (1976). Pattern recognition by means of disjoint principal components models. *Pattern recognition, 8*(3), 127-139.

Wold, S. (1978). Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics, 20*(4), 397-405.

Wold, S., Eriksson, L., Trygg, J., & Kettaneh, N. (2004). The PLS method–partial least squares projections to latent structures–and its applications in industrial RDP (research, development, and production). *Unea University*.

Wold, S., Ruhe, A., Wold, H., & Dunn, I., WJ. (1984). The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing, 5*(3), 735-743.

Wold, S., & Sjostrom, M. (1977). SIMCA: a method for analyzing chemical data in terms of similarity and analogy. *Chemometrics: theory and application, 52*, 243-282.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems, 58*(2), 109-130.

Woodcock, T., Downey, G., & O'Donnell, C. P. (2008). Confirmation of declared provenance of European extra virgin olive oil samples by NIR spectroscopy. *Journal of agricultural and food chemistry, 56*(23), 11520-11525.

Woodward, S., Waghorn, G., Attwood, G., & Li, D. (2010). *Ryegrass to lucerne-effects of dietary change on intake, milk yield and rumen microflora bacteria of dairy cows.* Paper presented at the Proceedings of the New Zealand Society of Animal Production.

Workman, J., Jr., & Weyer, L. (2007). *Practical Guide to Interpretive Near-Infrared Spectroscopy* (First ed.). Boca Raton: CRC Press.

Workman, J., Jr., & Weyer, L. (2008). *Practical Guide to Interpretive Near-Infrared Spectroscopy*: CRC press.

Workman, J., Jr., & Weyer, L. (2012). *Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy*. Boca Raton: CRC Press.

World Food Programme. (2014). Technical Specifications for Dried Whole Milk: Version V14.1. Retrieved from https://documents.wfp.org/stellent/groups/public/documents/manual_guide_proced/wfp259983.pdf

World Health Organisation. (2019). Food safety. Retrieved from https://www.who.int/news-room/fact-sheets/detail/food-safety

Wu, T., Chen, H., Lin, Z., & Tan, C. (2016). Identification and quantitation of melamine in milk by near-infrared spectroscopy and chemometrics. *Journal of Spectroscopy, 2016*.

Wu, W., Walczak, B., Massart, D., Heuerding, S., Erni, F., Last, I., & Prebble, K. (1996). Artificial neural networks in classification of NIR spectral data: design of the training set. *Chemometrics and Intelligent Laboratory Systems, 33*(1), 35-46.

Xiu, C., & Klein, K. (2010). Melamine in milk products in China: Examining the factors that led to deliberate use of the contaminant. *Food policy, 35*(5), 463-470.

Xu, L., Fu, X.-S., Cai, C.-B., & She, Y.-B. (2015). The feasibility of using near infrared spectroscopy for rapid discrimination of aged shiitake mushroom (Lentinula edodes) after long-term storage. *Journal of Chemistry, 2015*.

Xu, L., Shi, P.-T., Ye, Z.-H., Yan, S.-M., & Yu, X.-P. (2013). Rapid analysis of adulterations in Chinese lotus root powder (LRP) by near-infrared (NIR) spectroscopy coupled with chemometric class modeling techniques. *Food chemistry, 141*(3), 2434-2439.

Xu, L., Shi, W., Cai, C.-B., Zhong, W., & Tu, K. (2015). Rapid and nondestructive detection of multiple adulterants in kudzu starch by near infrared (NIR) spectroscopy and chemometrics. *LWT-Food Science and Technology, 61*(2), 590-595.

Zamora, R., & Hidalgo, F. J. (2011). The Maillard reaction and lipid oxidation. *Lipid Technology, 23*(3), 59-62.

Zhang, L.-G., Zhang, X., Ni, L.-J., Xue, Z.-B., Gu, X., & Huang, S.-X. (2014). Rapid identification of adulterated cow milk by non-linear pattern recognition methods based on near infrared spectroscopy. *Food chemistry, 145*, 342-348.

Zimmermann, B., & Kohler, A. (2013). Optimizing Savitzky-Golay parameters for improving spectral resolution and quantification in infrared spectroscopy. *Applied spectroscopy, 67*(8), 892-902.

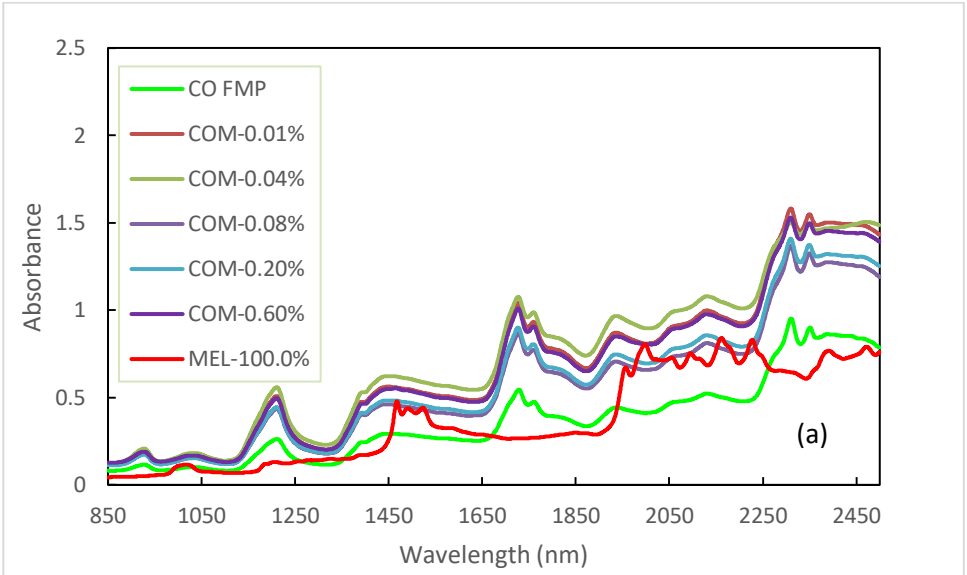
Appendix A
Effects of the addition of adulterants in FMP types in Chapter 5

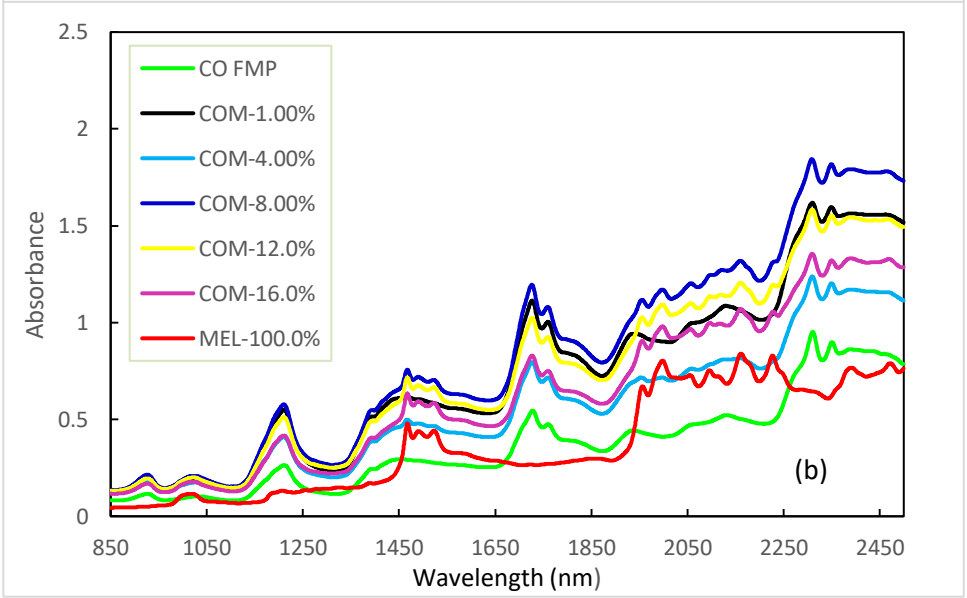
A.1 Changes in FMP crude protein contents

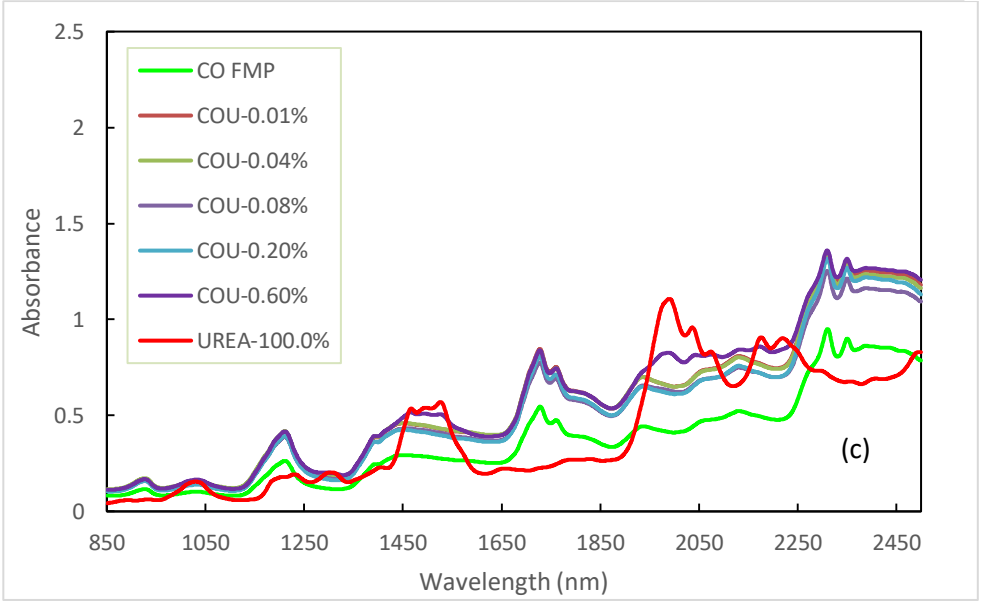
Table **A1:** Changes in crude protein contents (%) of the fat-filled milk powder (FMP) samples after the addition of 1 % of the chemical adulterants (i.e., melamine or urea)

FMP	Crude protein with melamine addition			Crude protein with urea addition		
	Without melamine	With melamine	Change (%)	Without urea	With urea	Change (%)
CO	7.94	11.59	45.98	7.94	11.35	43.01
PO	7.96	11.64	46.23	7.96	11.40	43.19
SBO	7.92	11.56	46.01	7.92	11.32	43.03
SFO	7.95	11.61	46.07	7.95	11.38	43.10

A.2 NIRS profiles of unadulterated and adulterated CO FMP type







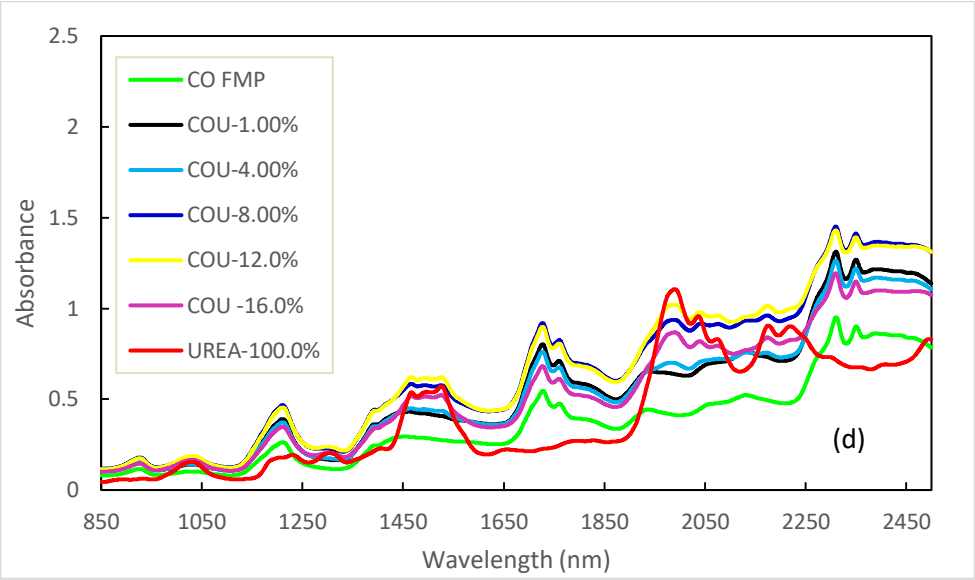
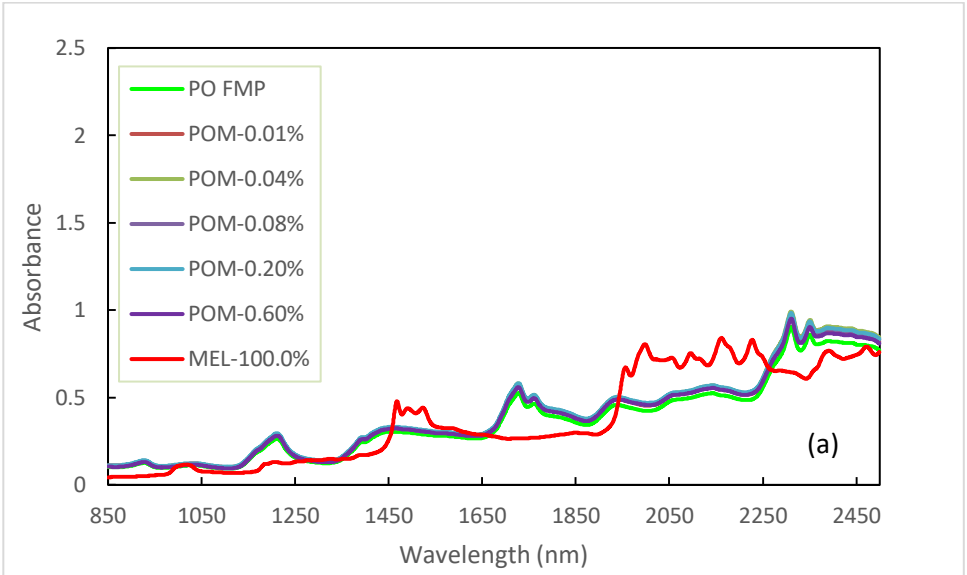
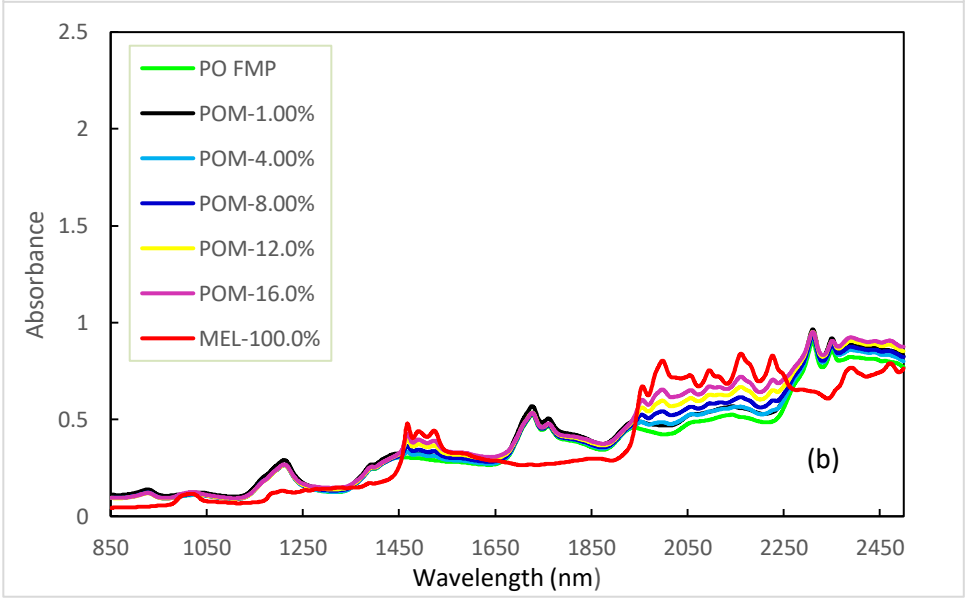
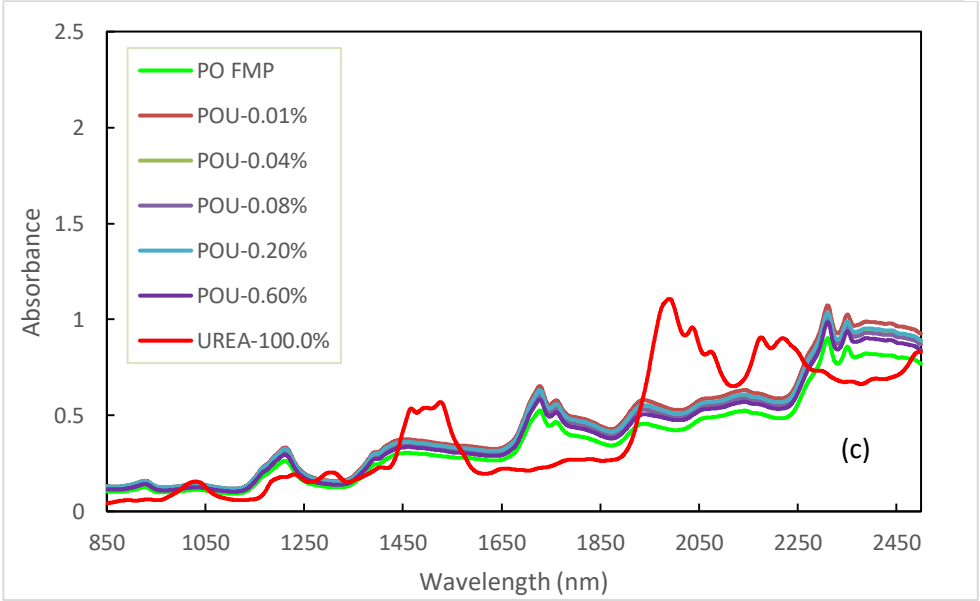


Figure **A2:** Representative spectra profiles of unadulterated and adulterated coconut oil FMP type, in addition to that of the pure (100%) adulterants. CO, COM, and COU represent the FMP types containing coconut oil, coconut oil with melamine, and coconut oil with urea; whereas MEL-100% and UREA-100% represent pure melamine and urea adulterants respectively. Graphs show FMP adulteration containing: (a) 0.01% - 0.60% melamine; (b) 1.00% - 16.0% melamine; (c) 0.01% - 0.60% urea; and (d) 1.00% - 16.0% urea.

A.3 NIRS profiles of unadulterated and adulterated PO FMP type







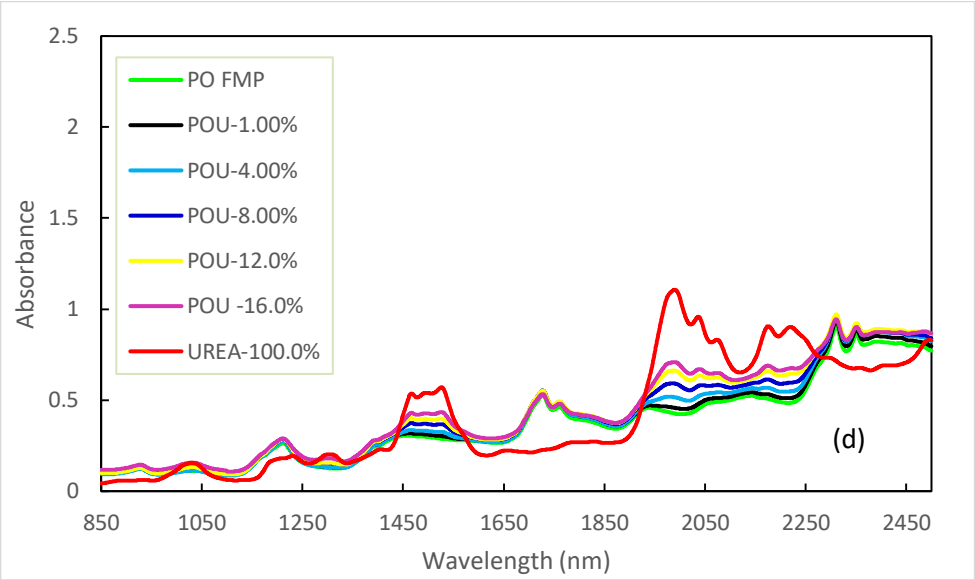
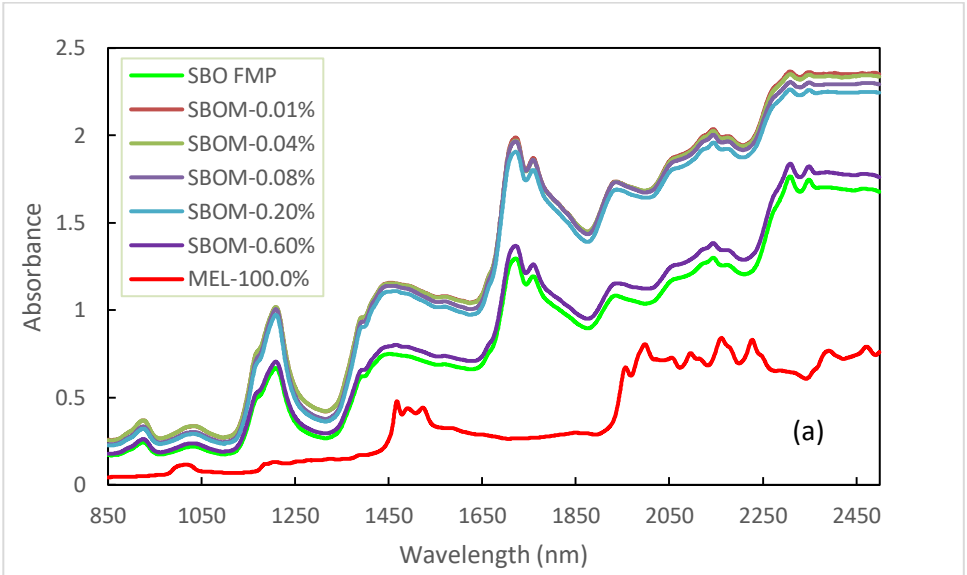
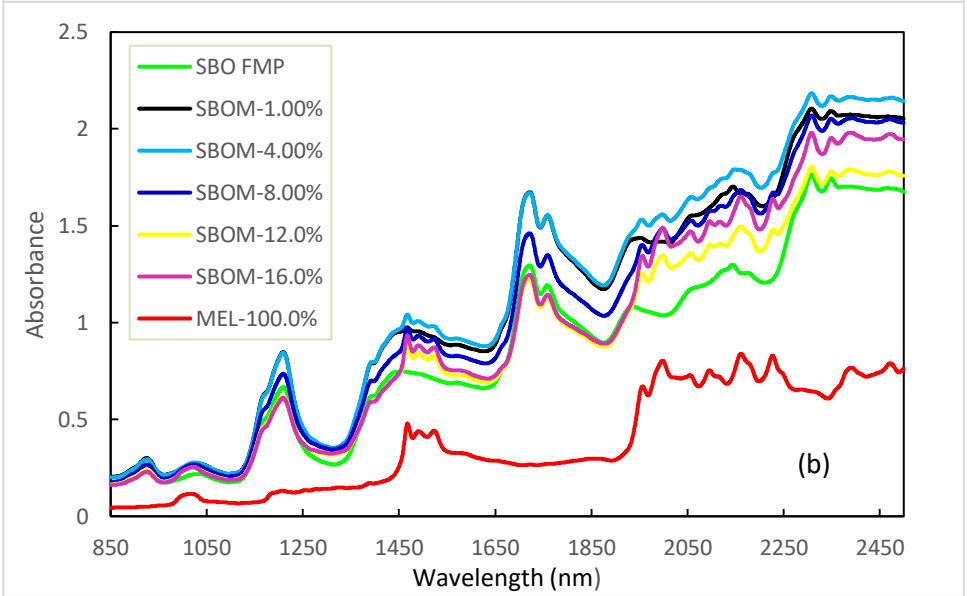
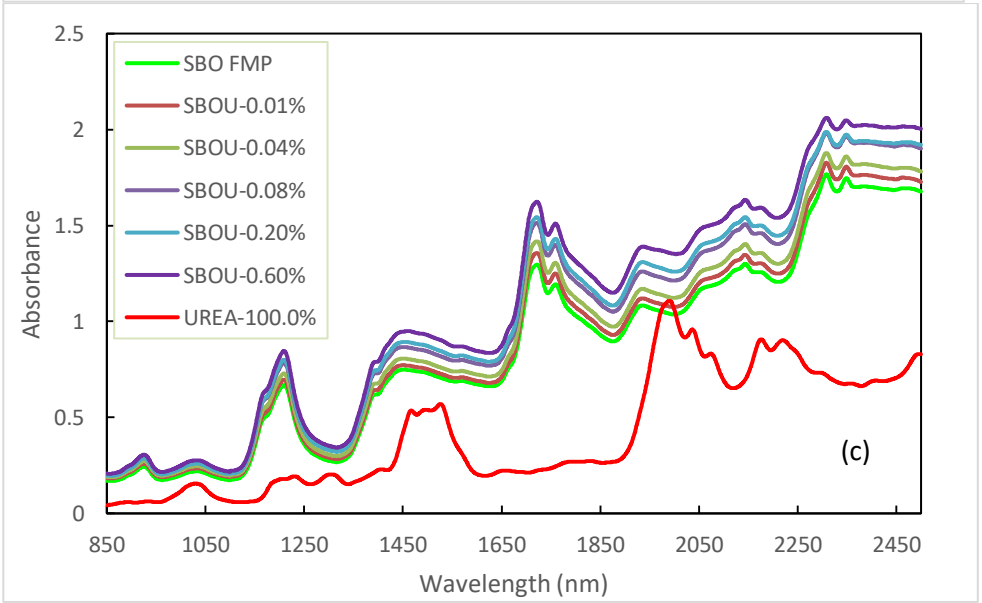


Figure **A3:** Representative spectra profiles of unadulterated and adulterated palm oil FMP type, in addition to that of the pure (100%) adulterants. PO, POM, and POU represent the FMP types containing palm oil, palm oil with melamine, and palm oil with urea; whereas MEL-100% and UREA-100% represent pure melamine and urea adulterants respectively. Graphs show FMP adulteration containing: (a) 0.01% - 0.60% melamine; (b) 1.00% - 16.0% melamine; (c) 0.01% - 0.60% urea; and (d) 1.00% - 16.0% urea.

A.4 NIRS profiles of unadulterated and adulterated SBO FMP type







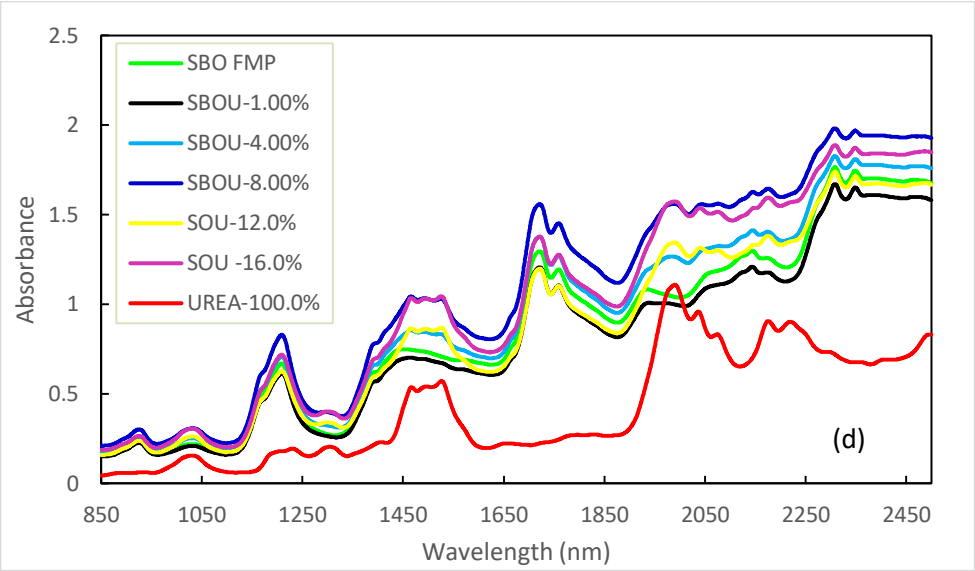
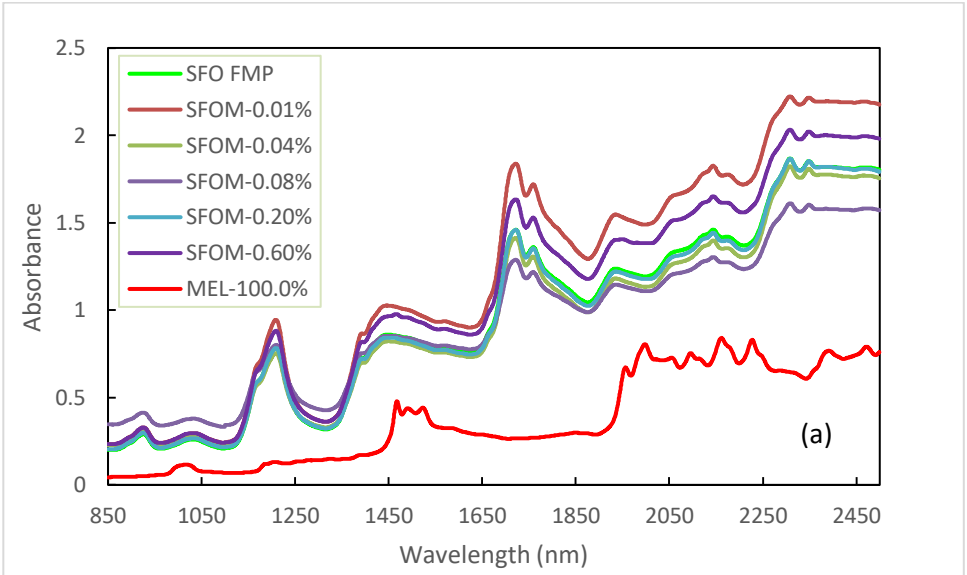
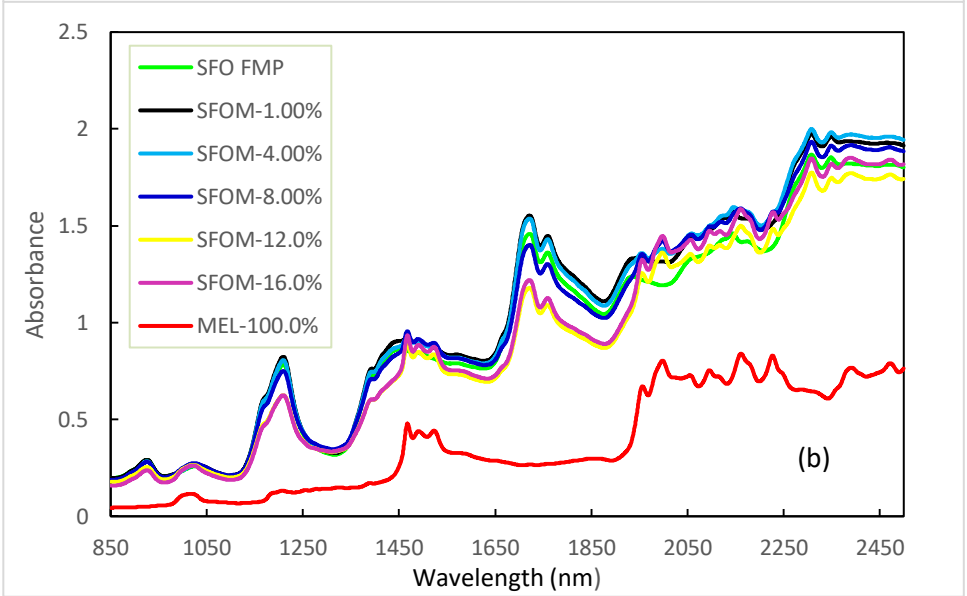
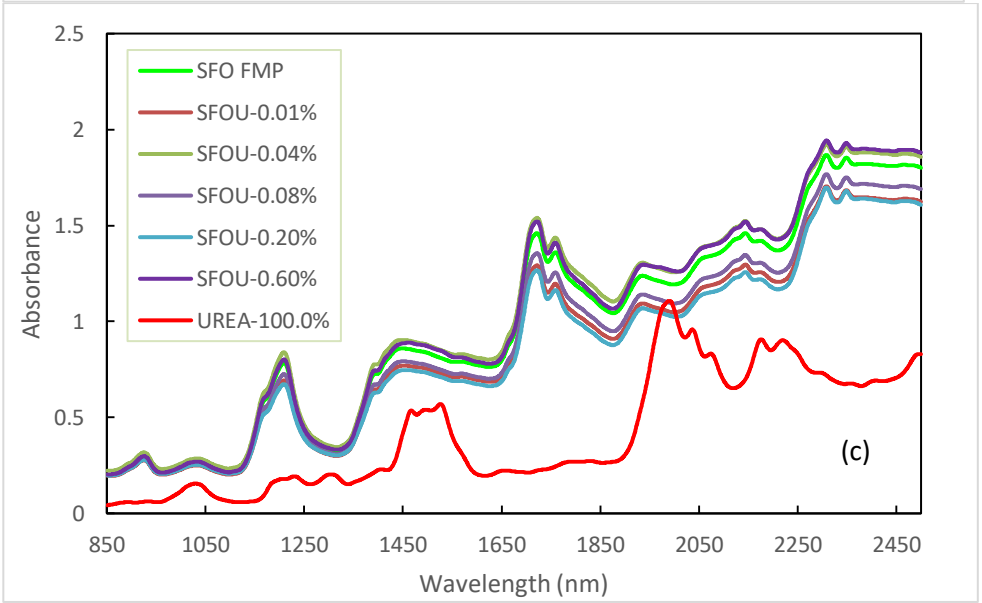


Figure **A4:** Representative spectra profiles of unadulterated and adulterated soya-bean oil FMP type, in addition to that of the pure (100%) adulterants. SBO, SBOM, and SBOU represent the FMP types containing soya-bean oil, soya-bean oil with melamine, and soya-bean oil with urea; whereas MEL-100% and UREA-100% represent pure melamine and urea adulterants respectively. Graphs show FMP adulteration containing: (a) 0.01% - 0.60% melamine; (b) 1.00% - 16.0% melamine; (c) 0.01% - 0.60% urea; and (d) 1.00% - 16.0% urea.

A.5 NIRS profiles of unadulterated and adulterated SFO FMP type







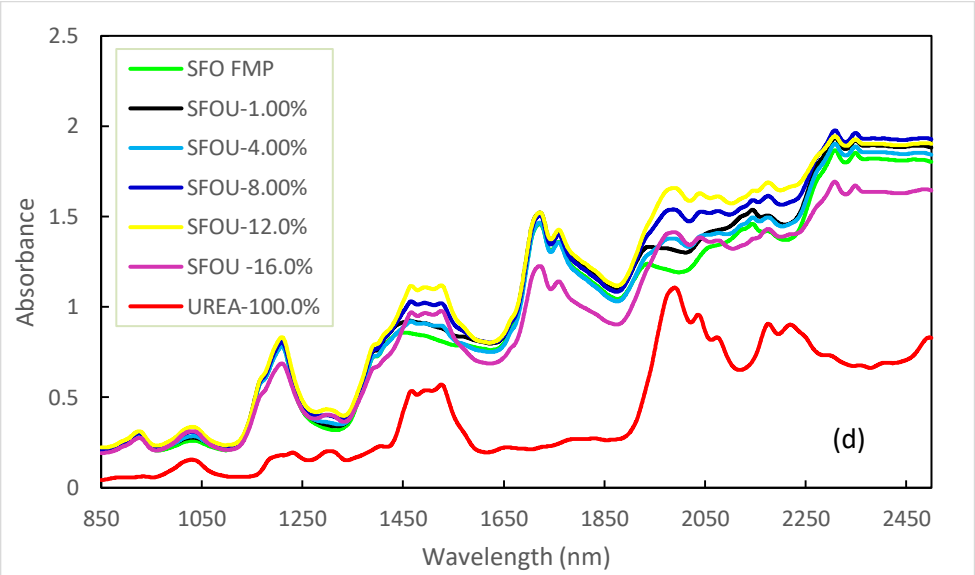


Figure **A5:** Representative spectra profiles of unadulterated and adulterated sunflower oil FMP type, in addition to that of the pure (100%) adulterants. SFO, SFOM, and SFOU represent the FMP types containing sunflower oil, sunflower oil with melamine, and sunflower oil with urea; whereas MEL-100% and UREA-100% represent pure melamine and urea adulterants respectively. Graphs show FMP adulteration containing: (a) 0.01% - 0.60% melamine; (b) 1.00% - 16.0% melamine; (c) 0.01% - 0.60% urea; and (d) 1.00% - 16.0% urea.

A.6 Description of the algorithms used in Chapter 5

The descriptions of the algorithms used for the performance evaluation of the classification and the quantification models for the detection of the chemical adulterants in the FMP types are given below.

The efficiency (E) of the model was calculated as the geometric mean of the values of sensitivity and specificity (Chong & Jun, 2005; Oliveri & Downey, 2012):

$$E = \sqrt{Sensitivity \times Specificity} \tag{5.1}$$

The Reliability rate (RR) was calculated according to Reis et al. (2017) to provide a better overview of the trueness of the models and it is given as:

$$RR = Sensitivity + Specificity - 1 \tag{5.2}$$

RMSEP which is an estimate of how close the measured data points are to the predicted values of the *n* independent validation set not contained in the calibration set (Porep et al., 2015) was calculated using:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(P_i - M_i)^2}{n}} \tag{5.3}$$

However, RSR which standardises RMSEP using the observations standard deviation (STDEV) was calculated to provide a delimiter of what is considered a low RMSEP value (Moriasi et al., 2007). Hence, for the *n* samples with M measured mean values,

$$RSR = \frac{RMSEP}{STDEV} = \sqrt{\frac{(n-1)}{n} \times \frac{\sum_{i=1}^{n}(P_i - M_i)^2}{\sum_{i=1}^{n}(M_i - M)^2}} \tag{5.4}$$

RSR varies from an optimal value of 0, which indicates zero RMSEP and therefore perfect model prediction, to a large positive value (Moriasi et al., 2007).

PBIAS, expressed as a percentage, was calculated to measure the average deviation of the predicted data from the measured values (Moriasi et al., 2007) and is given as:

$$PBIAS = \frac{\sum_{i=1}^{n}(M_i - P_i) \times 100}{\sum_{i=1}^{n}(M_i)} \quad (5.5)$$

The prediction sum of squares (PRESS) R^2 was calculated using:

$$R_P^2 = 1 - \frac{\sum_{i=1}^{n}(P_i - M_i)^2}{\sum_{i=1}^{n}(M_i - M)^2} \quad (5.6)$$

Appendix B

Changes in FA and AA FMP types after storage in Chapter 6

B.1 Paired t-test of the changes in FA of the FMP types

Table **B1:** Paired t-test of the differences in fatty acids (FA) contents of the four fat-filled milk powder types formulated at 0 week (4 °C) and stored at 40 °C for 7 weeks

Fatty Acids	Mean change	Standard Deviation	t-Value	Df	p-Value
C6:0	-0.01	0.01	-1.00	3	0.391
C8:0	0.13	0.25	1.06	3	0.366
C10:0	0.02	0.03	1.34	3	0.272
C12:0	0.24	0.50	0.97	3	0.403
C14:0	-0.012	0.10	-0.22	3	0.837
C16:0	-0.13	0.37	-0.70	3	0.532
C18:0	-0.34*	0.071	-9.70	3	0.002
∑SFA	-0.09	0.65	-0.28	3	0.795
C18:1 c9	0.32	0.44	1.48	3	0.235
C18:2 c9,12	0.02	0.25	0.16	3	0.882
C18:3c9,12,15	0.18	0.36	1.00	3	0.393
∑UFA	0.52	0.55	1.88	3	0.157
OA/LA	0.04	0.07	1.15	3	0.332
∑SFA/∑UFA	0.10	0.23	0.91	3	0.431
MUFA	0.32	0.44	1.48	3	0.235
PUFA	0.20	0.49	0.81	3	0.479
IA	0.15	0.31	0.95	3	0.413
IT	-0.01	0.04	-0.14	3	0.895

*indicates the mean change (expressed as g/100 g of FA) of the FA contents that is statistically significant at $p < 0.05$. Df means the degree of freedom.

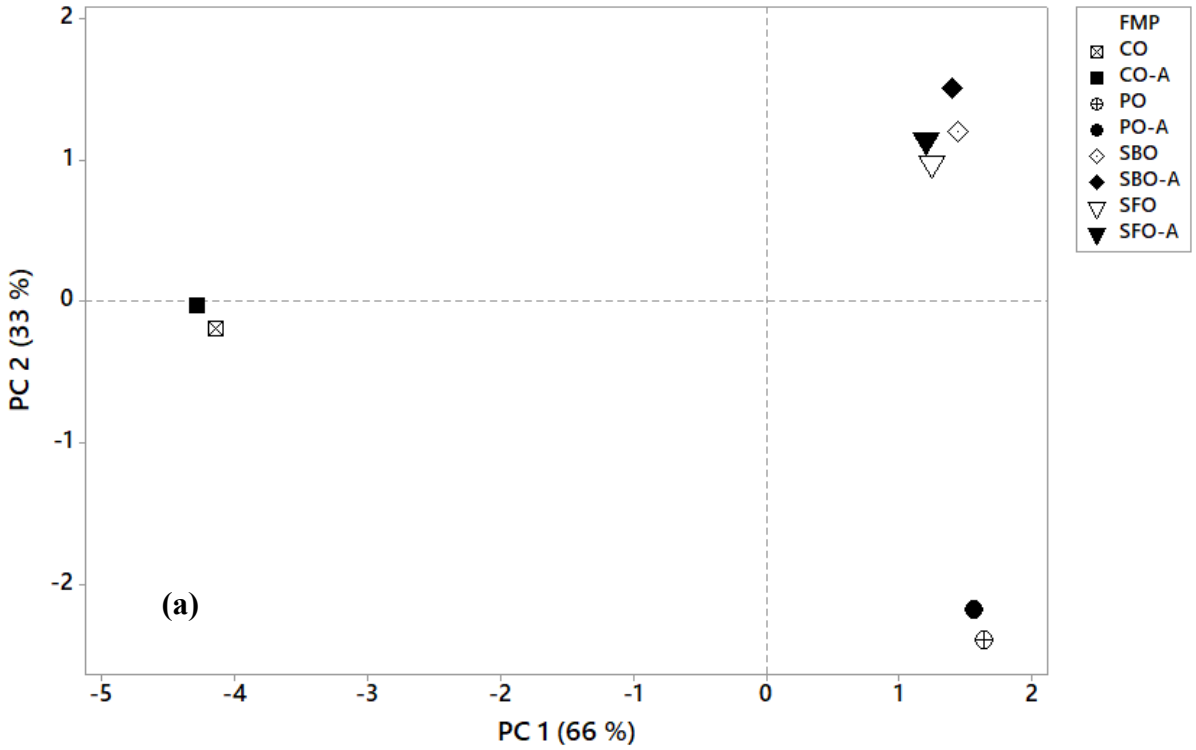
B.2 Paired t-test of the changes in AA of the FMP types

Table **B2:** Paired t-test of the differences in amino acids contents of the four fat-filled milk powder types formulated at 0 week (4 °C) and stored at 40 °C for 7 weeks

Amino Acids	Mean change	Standard Deviation	t-Value	Df	p-Value
Lys	-0.47*	0.06	-16.47	3	0.000
Thr	-0.12	0.08	-3.08	3	0.054
Trp	-0.03*	0.01	-4.24	3	0.024
Leu	0.20*	0.06	6.39	3	0.008
Ile	-0.43*	0.06	-15.21	3	0.001
Val	-0.05	0.03	-2.63	3	0.078
Met	-0.09*	0.03	-5.67	3	0.011
Phe	-0.09*	0.03	-5.44	3	0.012
His	0.018	0.04	0.87	3	0.449
∑EAA	-1.06*	0.28	-7.52	3	0.005
Arg	-0.26	0.19	-2.80	3	0.068
Pro	-0.07	0.07	-2.02	3	0.137
Cys	0.01	0.01	1.57	3	0.215
Gly	-0.08*	0.03	-4.82	3	0.017
∑CEAA	-0.40	0.24	-3.32	3	0.045
Ala	-0.10*	0.04	-5.62	3	0.011
Asp	-0.20*	0.04	-10.22	3	0.002
Glu	0.033	0.12	0.53	3	0.633
Ser	-0.13*	0.05	-5.46	3	0.012
Tyr	-0.07*	0.03	-5.17	3	0.014
∑NEAA	-0.46*	0.26	-3.63	3	0.036

*indicates the mean changes (expressed as g/100 g of milk) of the amino acids contents that are statistically significant at $p < 0.05$. Df means the degree of freedom.

B.3 PCA of the FA profiles of the fresh and aged FMP types


FMP
CO
CO-A
PO
PO-A
SBO
SBO-A
SFO
SFO-A
PC 2 (33 %)
PC 1 (66 %)
2
1
0
-1
-2
-5
-4
-3
-2
-1
0
1
2
(a)

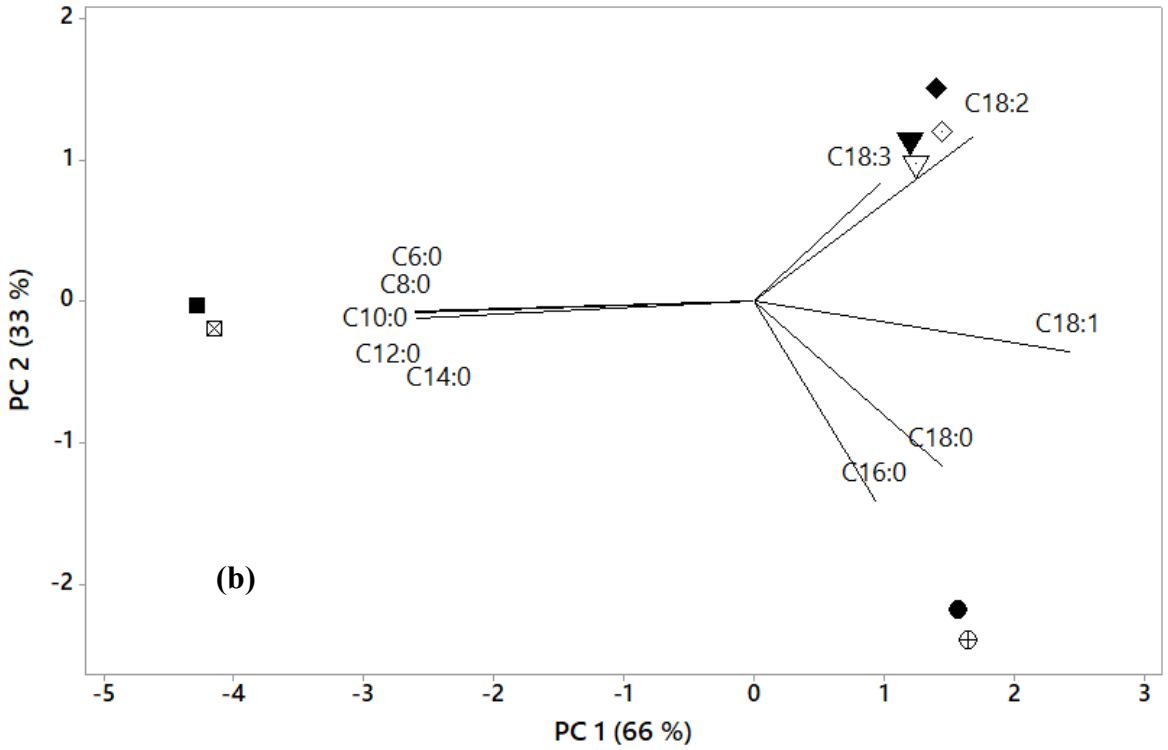
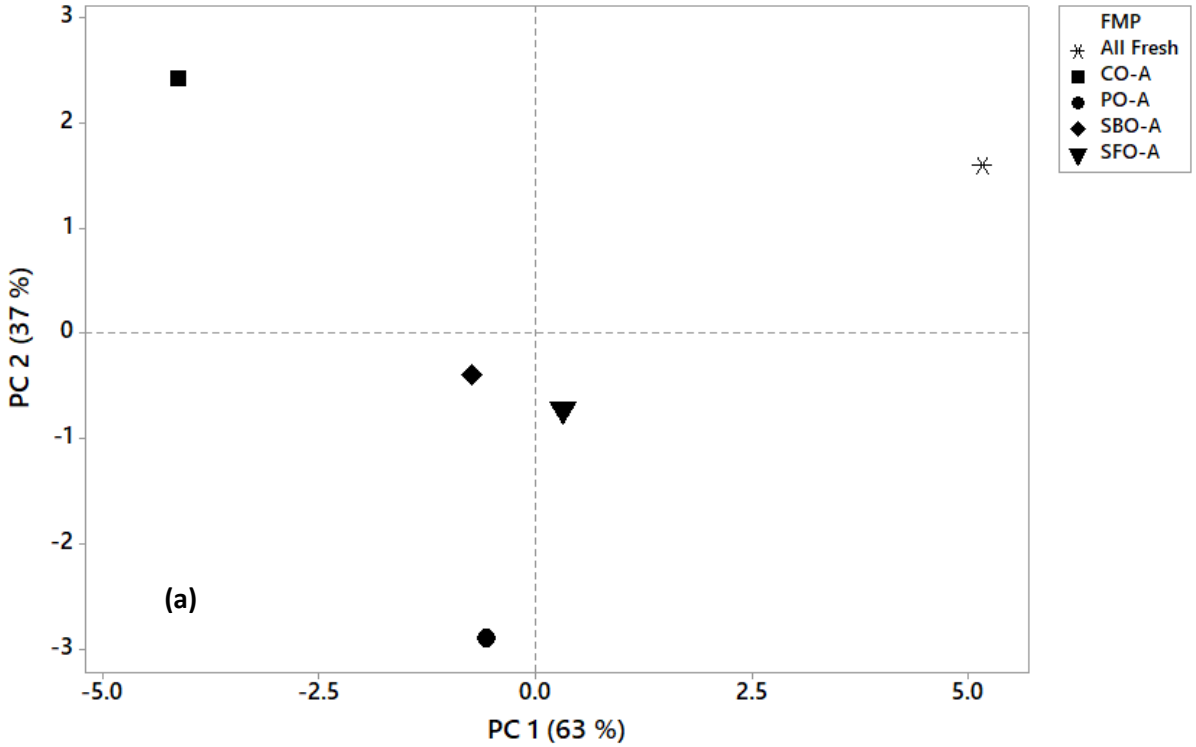


Figure **B3:** PCA (a) score plot and (b) biplot of the fatty acid profiles of the fresh (CO, PO, SBO AND SFO) and the aged (CO-A, PO-A, SBO-A and SFO-A) FMP types. CO, PO, SBO and SFO represent FMP types containing coconut oil, palm oil, soya-bean oil and sunflower oil.

B.4 PCA of the AA profiles of the fresh and aged FMP types



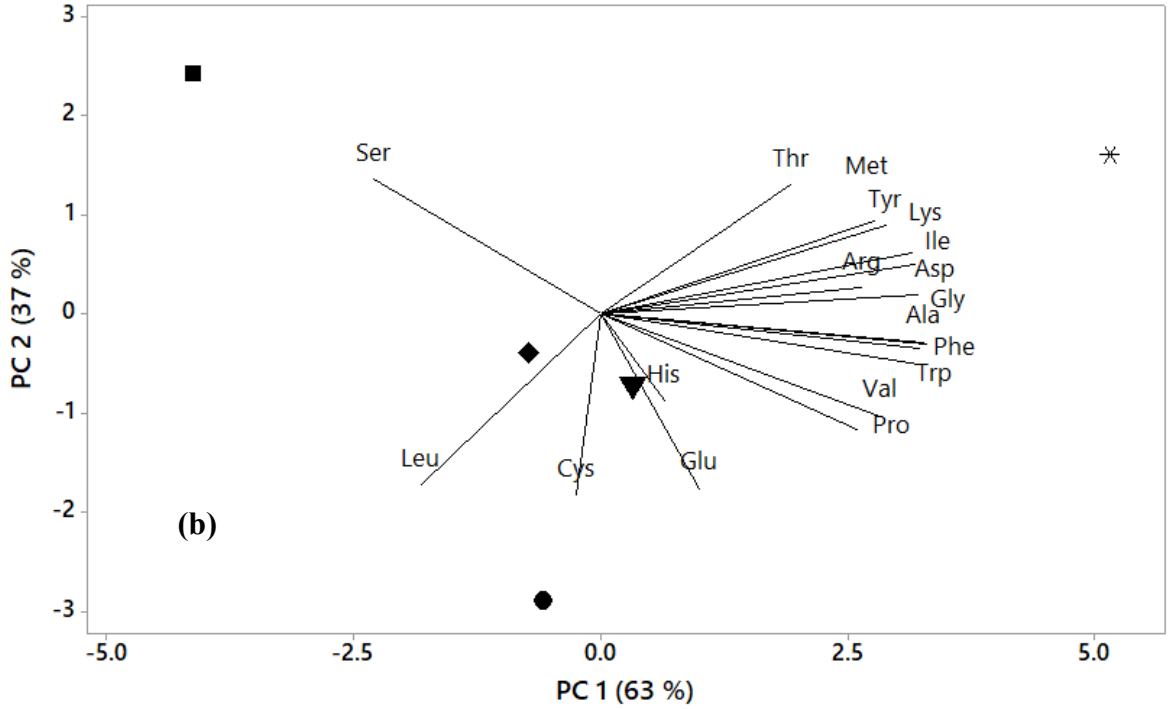


Figure **B4:** PCA (a) score plot and (b) biplot of the amino acid profiles of the fresh (CO, PO, SBO and SFO) and the aged (CO-A, PO-A, SBO-A and SFO-A) FMP types. CO, PO, SBO and SFO represent FMP types containing coconut oil, palm oil, soya-bean oil and sunflower oil. All fresh FMP types had the same amount of AA at t = 0 week.

Appendix C

Statistics and NIR spectra comparisons of FMP brands in Chapter 7

C.1 Paired t-test of the changes in FA of the Nigerian FMP brands

Table **C1**: Paired t-test of the differences in fatty acids (FA) contents of the seven Nigerian fat-filled milk powder brands received at 0 week (4 °C) and stored at 40 °C for 7 weeks

Fatty Acids	Mean change	Standard Deviation	t-Value	Df	p-Value
C12:0	0.006	0.011	1.49	6	0.186
C14:0	-0.009	0.031	-0.74	6	0.490
C16:0	-0.059	0.108	-1.44	6	0.200
C18:0	-0.007	0.050	-0.37	6	0.724
C20:0	-0.002	0.003	-1.62	6	0.156
∑SFA	-0.070	0.081	-2.31	6	0.060
C16:1 c9	0.000	0.003	-0.01	6	0.996
C18:1 c9	0.100	0.187	1.41	6	0.208
C18:2 c9,12	-0.023	0.135	-0.45	6	0.667
C18:3c9,12,15	-0.003	0.007	-1.28	6	0.248
∑UFA	0.073	0.093	2.09	6	0.082
LA/ALA	0.312	0.558	1.48	6	0.189
IA	-0.004	0.005	-2.17	6	0.073
IT	-0.006	0.007	-2.18	6	0.072

*indicates the mean change (expressed as g/100 g of FA) of the FA contents that is statistically significant at $p < 0.05$. Df means the degree of freedom.

C.2 Paired t-test of the changes in AA of the Nigerian FMP brands

Table **C2:** Paired t-test of the differences in amino acids contents of the seven Nigerian fat-filled milk powder brands received at 0 week (4 °C) and stored at 40 °C for 7 weeks

Amino Acids	Mean change	Standard Deviation	t-Value	Df	p-Value
Lys	-0.085*	0.065	-3.48	6	0.013
Thr	-0.008	0.019	-1.07	6	0.328
Trp	-0.034*	0.018	-5.07	6	0.002
Leu	-0.047	0.051	-2.44	6	0.050
Ile	-0.014	0.032	-1.16	6	0.291
Val	-0.010	0.034	-0.79	6	0.456
Met	0.002	0.025	0.23	6	0.824
Phe	-0.001	0.031	-0.10	6	0.927
His	0.058	0.087	1.76	6	0.129
Arg	-0.002	0.067	-0.09	6	0.933
Pro	-0.031	0.134	-0.61	6	0.562
Cys	-0.005	0.022	-0.54	6	0.606
Gly	0.004	0.008	1.24	6	0.262
Tau	0.000	0.001	-0.84	6	0.431
∑EAA	-0.174	0.326	-1.41	6	0.208
∑AA	-0.452*	0.467	-2.56	6	0.043

*indicates the mean changes (expressed as g/100 g of milk) of the amino acids contents that are statistically significant at $p < 0.05$. Df means the degree of freedom.

C.3 SIMCA/PLSR models of FMP brands based on Kennard-Stone algorithm

Table **C3:** Performance statistics of the multiclass SIMCA and PLSR models for Nigerian fat-filled milk powders (FMP) received at t = 0 week and stored for 7 weeks at 40 °C and 90 % relative humidity

SIMCA classification[1] of fresh (t = 0 week) Nigerian FMP brands							
	Sensitivity and specificity of fresh test sets (n = 35) on assignation into classes						
True class[2]	3-Crown	BlueBoat	Cowbell	Dano	Luna	Olympic	Peak
3-Crown	100.0	100.0	100.0	100.0	100.0	100.0	100.0
BlueBoat	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Cowbell	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Dano	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Luna	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Olympic	100.0	100.0	100.0	100.0	100.0	80.0	100.0
Peak	100.0	100.0	100.0	100.0	100.0	100.0	100.0
SIMCA classification of aged (t = 7 weeks) Nigerian FMP brands							
	Specificity values of all aged samples (n = 147) on assignation into classes						
True class	3-Crown	BlueBoat	Cowbell	Dano	Luna	Olympic	Peak
3-Crown	100.0	100.0	100.0	100.0	100.0	100.0	100.0
BlueBoat	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Cowbell	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Dano	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Luna	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Olympic	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Peak	100.0	100.0	100.0	100.0	100.0	100.0	100.0
PLSR predictions[3] of the storage time in weeks of Nigerian FMP brands							
	FMP brands aged/stored at 40 °C for 7 weeks						
Metrics[4]	3-Crown	BlueBoat	Cowbell	Dano	Luna	Olympic	Peak
R^2p	0.75	0.99	0.99	0.99	0.88	0.85	0.97
RMSECV	0.58	0.20	0.27	0.24	0.25	0.60	0.48
RMSEP	1.10	0.28	0.19	0.21	1.05	1.00	0.48
RPD	1.78	5.69	8.15	11.28	1.83	1.68	2.57
RSR	0.57	0.18	0.12	0.09	0.55	0.60	0.39
NSE	0.59	0.96	0.98	0.99	0.63	0.56	0.81

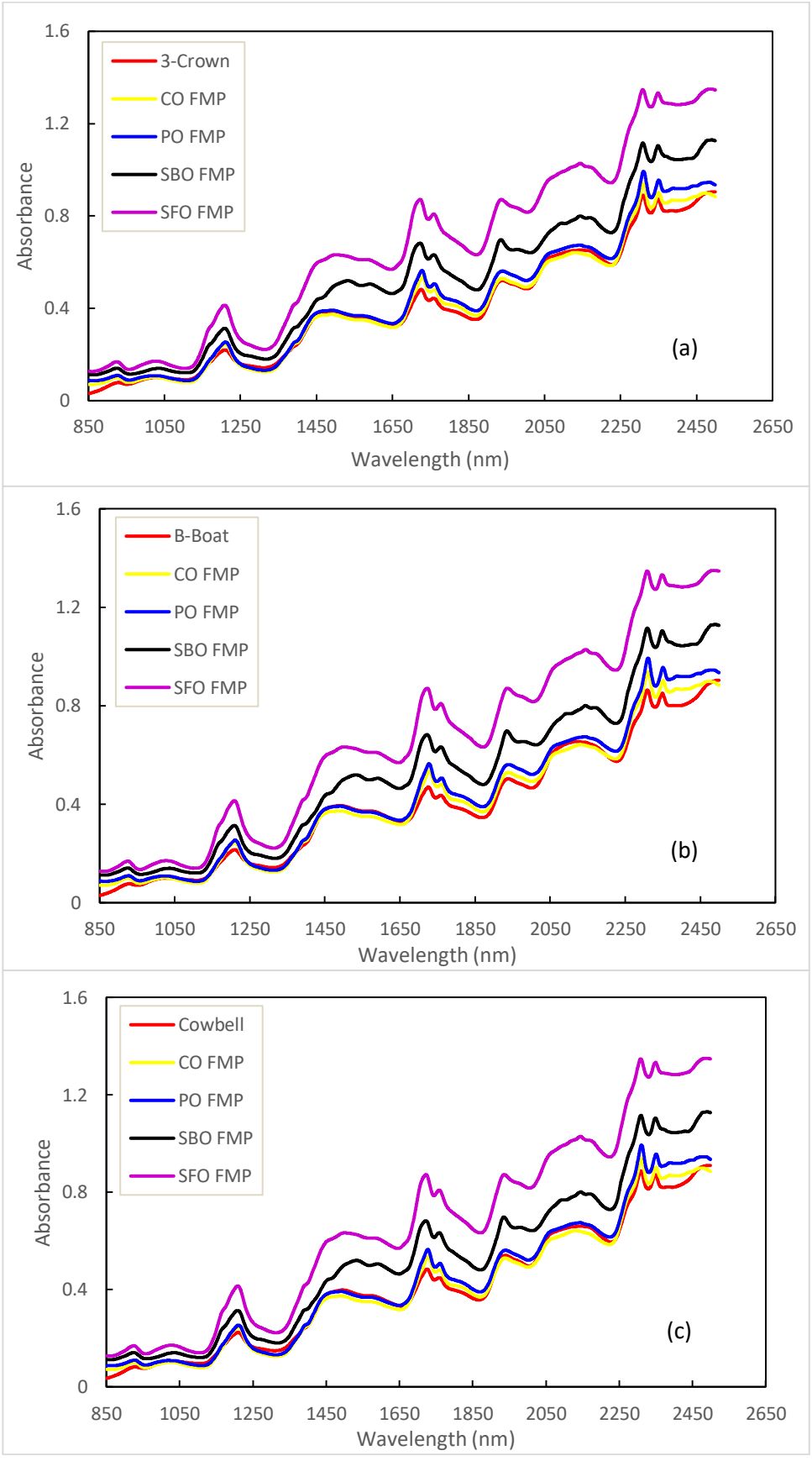
[1]Classification models were built using only the sensitive bands, 1725 – 2050 nm, of the preprocessed calibration sets (n = 105; i.e., 15 from each of the 7 fresh brands) selected based on Kennard-Stone algorithm.

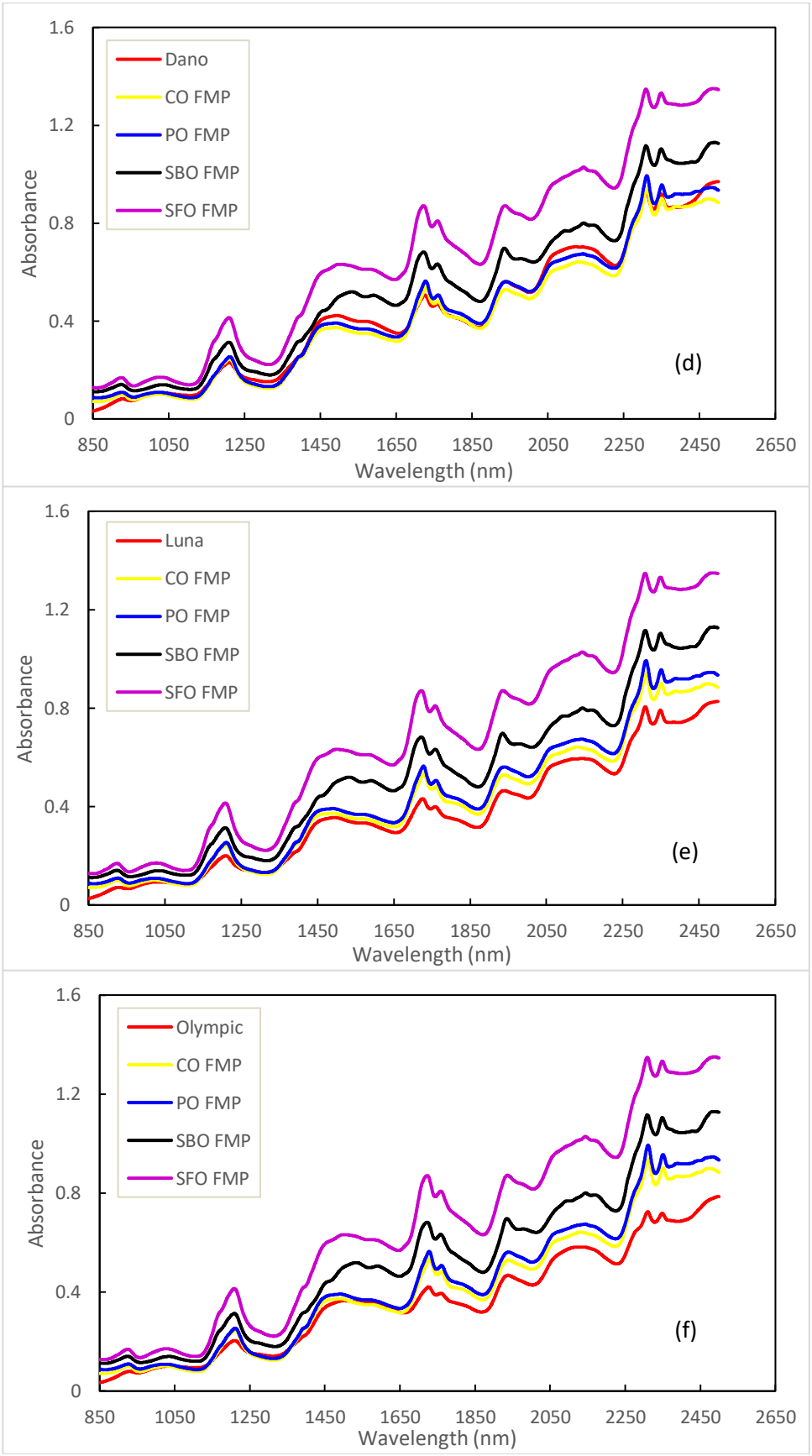
[2]True class represents the one-class model built for each of the brands using the calibration sets of the fresh samples at t = 0 week.

[3]Prediction models were built using only the sensitive intervals (i.e., interval PLS) of the preprocessed calibration sets (n = 112; i.e., 16 from each of the 7 aged brands) selected based on Kennard-Stone algorithm.

[4]R^2p = coefficient of determination for prediction (n = 35); RMSECV = root mean square error of cross validation (weeks); RMSEP = root mean square error of prediction (weeks); RSR = RMSE-observation standard deviation ratio calculated as 1/RPD (ratio of prediction to deviation) (optimal RSR < 0.5); NSE = Nash-Sutcliffe efficiency.

C.4 Spectra comparisons of unadulterated FMP types with the FMP brands




Dano
CO FMP
PO FMP
SBO FMP
SFO FMP
Absorbance
Wavelength (nm)
(d)
Luna
CO FMP
PO FMP
SBO FMP
SFO FMP
Absorbance
Wavelength (nm)
(e)
Olympic
CO FMP
PO FMP
SBO FMP
SFO FMP
Absorbance
Wavelength (nm)
(f)

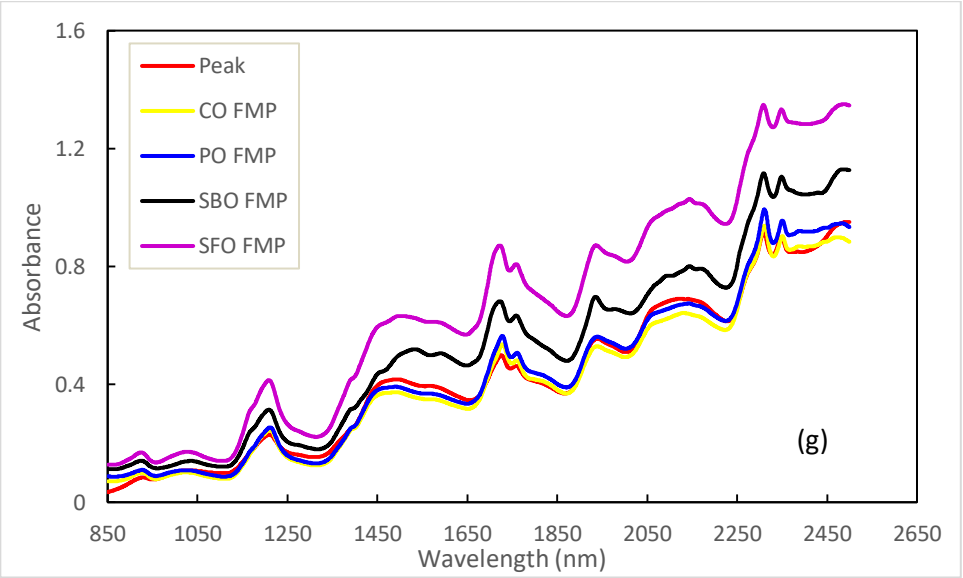
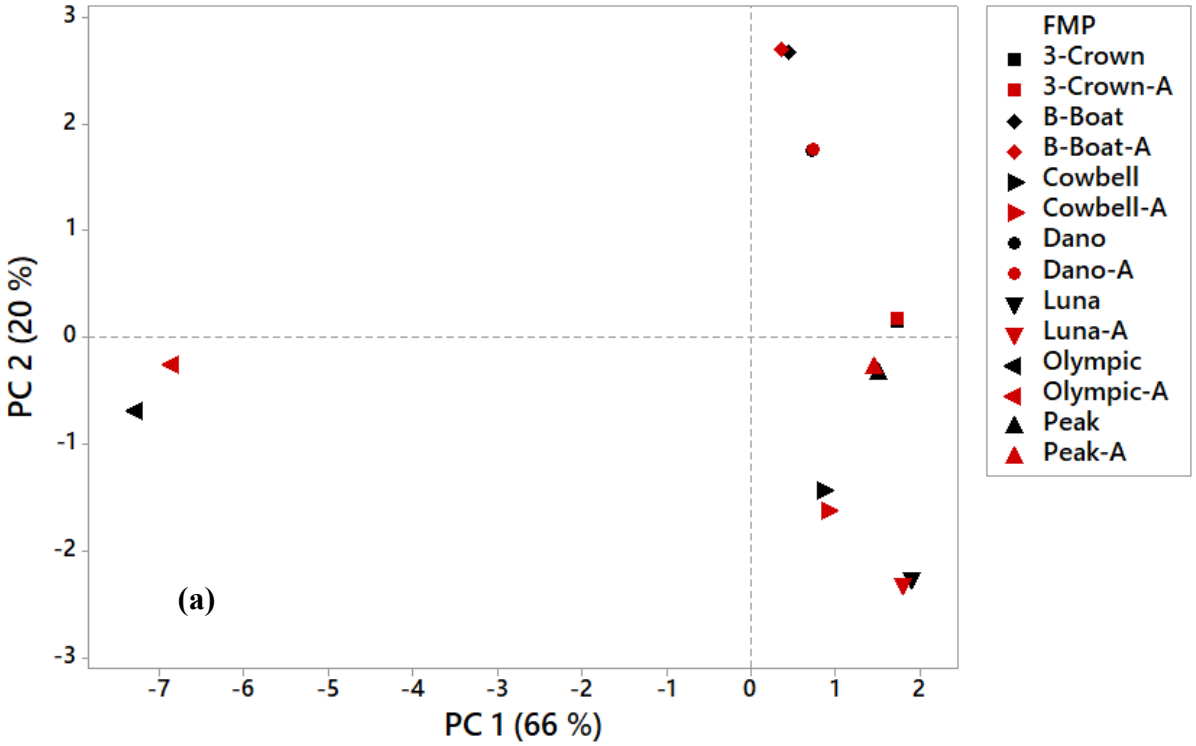


Figure **C4:** Representative near infrared spectra profiles of unadulterated fat-filled milk powder (FMP) types (i.e., CO (coconut oil), PO (palm oil), SBO (soya-bean oil) and SFO (sunflower oil)); and that of 7 commercial FMP brands (i.e., (a) 3-crown, (b) blue-boat, (c) cowbell, (d) dano, (e) luna, (f) olympic, and (g) peak) obtained from Nigerian markets.

C.5 PCA of the FA profiles of the fresh and aged Nigerian FMP brands



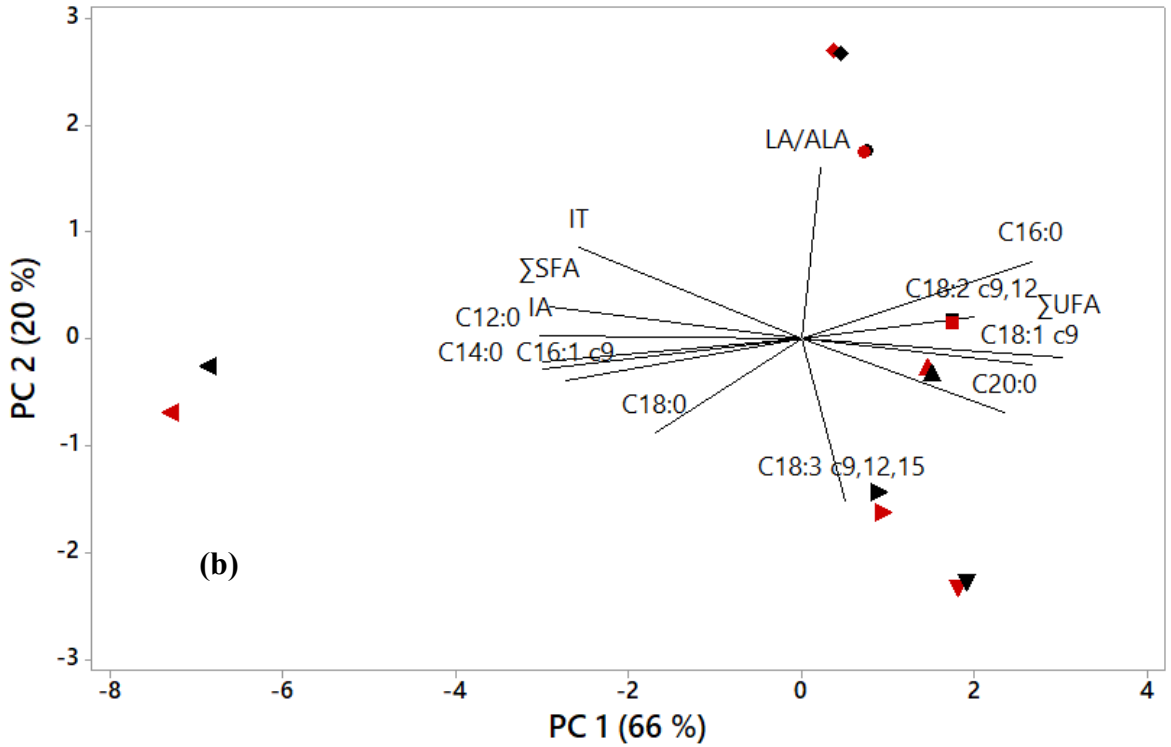
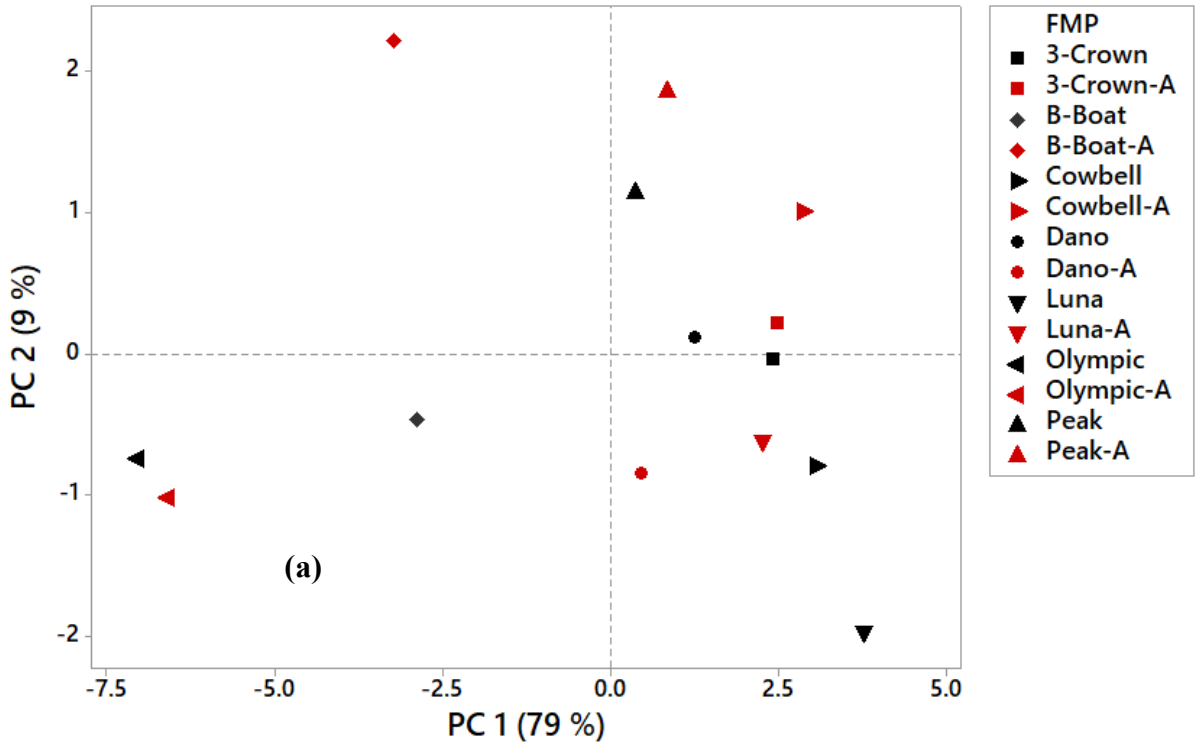


Figure **C5:** PCA (a) score plot and (b) biplot of the fatty acid profiles of the fresh (3-Crown, B-Boat, Cowbell, Dano, Luna, Olympic and Peak) and the aged (3-Crown-A, B-Boat-A, Cowbell-A, Dano-A, Luna-A, Olympic-A and Peak-A) Nigerian FMP brands.

C.6 PCA of the AA profiles of the fresh and aged Nigerian FMP brands



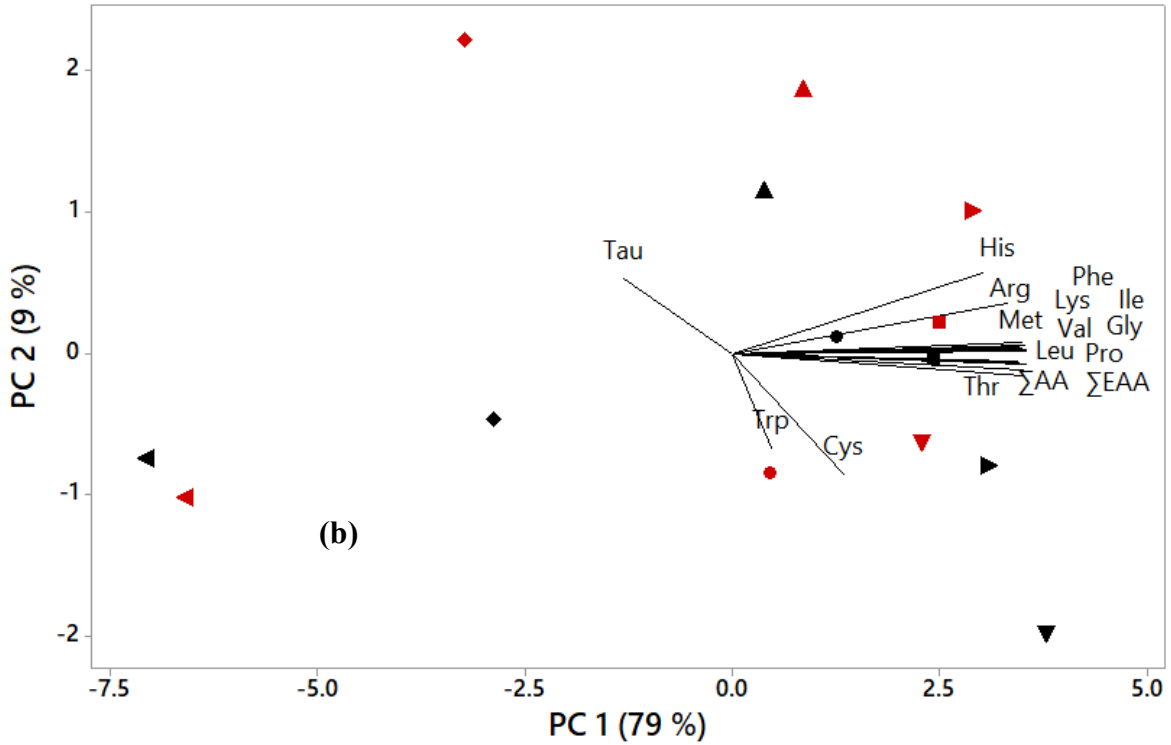


Figure **C6:** PCA (a) score plot and (b) biplot of the amino acid profiles of the fresh (3-Crown, B-Boat, Cowbell, Dano, Luna, Olympic and Peak) and the aged (3-Crown-A, B-Boat-A, Cowbell-A, Dano-A, Luna-A, Olympic-A and Peak-A) Nigerian FMP brands.